

ial

Issues in Applied Linguistics

Volume 12 Number 1 • June 2001

ARTICLES

Wh- Movement in Child Catalan

John Grinstead

Balancing the Competing Interests in Seminar Discussion: Peer Referencing and Asserting Vulnerability

Hansun Zhang Waring

Cohesion and Coherence in Children's Written English: Immersion and English-only Classes

Jungok Bae

INTERVIEWS

An Interview with J. Charles Alderson

Viphavee Vongpumivitch and Nathan Carr

An Interview with Dorry M. Kenyon

Nathan Carr and Viphavee Vongpumivitch



ial

Issues in Applied Linguistics

Volume 12 Number 1 • June 2001

ial

Issues in Applied Linguistics

Volume 12 Number 1 • June 2001

Editors

David Olsher
Leah Wingard

Assistant Editors

Debra Friedman
Emmy Goldknopf

Review Editors

Stefan Frazier
Mikyung Kim

Managing Editor

Viphavee Vongpumivitch

Production Editor and Webmaster

Scott Phillabaum

Treasurer

Nathan Carr

Proofreaders

Olga Griswold, Shimako Iwasaki, Anna Rodriguez

Additional Readers

Galina Bolden, Ignasi Clemente, Carleen Curley, Kamil Ud Deen, Sally Jacoby, Brian Lynch, James Purpura, Howard Williams, Lexie Wood

Issues in Applied Linguistics is published twice a year by the graduate students of the Department of Applied Linguistics & TESL at UCLA.

The views expressed in the journal are not necessarily those of the Editors, the UCLA Department of Applied Linguistics & TESL, or the Regents of the University of California.

IAL is printed at Lue's Printers: Los Angeles, California.

Partial funding for *IAL* is provided by the UCLA Graduate Students Association.

IAL is abstracted in *Linguistics and Learning Behavior Abstracts (LLBA)*, on *Comserve*, and is listed in the *MLA International Bibliography*.

IAL's European distributor is LINCOM EUROPA.

Copyright © 2001, Regents of the University of California ISSN 1050-4273

CONTENTS

ial

Editorial	
David Olsher	
Leah Wingard	
	1

ARTICLES

Wh- Movement in Child Catalan	
John Grinstead	5

Balancing the Competing Interests in Seminar Discussion: Peer Referencing and Asserting Vulnerability	
Hansun Zhang Waring	29

Cohesion and Coherence in Children’s Written English: Immersion and English-only Classes	
Jungok Bae	51

INTERVIEWS

Two Interviews on Language Testing: An Introduction	
Nathan Carr and Viphavee Vongpumivitch	89

An Interview with J. Charles Alderson	
Viphavee Vongpumivitch and Nathan Carr	91

An Interview with Dorry M. Kenyon	
Nathan Carr and Viphavee Vongpumivitch	111

BOOK REVIEWS

Language Testing

Tim McNamara

Reviewed by Kirby J. Cook

127

How Children Learn the Meanings of Words

Paul Bloom

Reviewed by Masahiko Minami

131

Memory: From Mind to Molecules

Larry Squire and Eric Kandel

Reviewed by Sheila Crowell

135

Publications Received

138

Subscription Information

140

Back Issue Information

141

Information for Contributors

144

Editorial

Issues in Applied Linguistics continues its commitment to publishing research representing diverse perspectives and research traditions in applied linguistics. This issue, which marks our twelfth year of production, reflects the breadth of our field by bringing together studies of language acquisition, discourse analysis, and language assessment. The articles reflect the interrelatedness of research across these core areas of applied linguistics. A study of child language acquisition from a formal linguistic perspective makes use of naturally occurring language data, reflecting the acquisition researcher's interest in data drawn from authentic discourse. A discourse analytic study of the speaking practices evident in graduate-level university seminars has implications for understanding the competencies needed by speakers, native or nonnative, functioning within such institutional contexts. A language assessment study aimed at defining constructs of coherence and cohesion that can be used to assess the writing of elementary school children draws on, and ultimately adds to, our understanding of the discourse analytic concepts of coherence and cohesion in written texts. It is not that the authors of these studies, or we as editors, set out to show such connections, but rather that these connections exist and are reflected in the crosscutting orientations and implications of research projects working from diverse goals and paradigms. That said, these articles also deserve to be considered on their own terms.

Working within a formal linguistic framework, John Grinstead examines the development of question formation in four Catalan children. He finds that the onset of adult-like Wh- question formation appears to correlate with the onset of production of a much wider variety of tense morphology. Grinstead suggests that children need to acquire contrastive tense features in order to make available a structural attachment site for the Wh- feature which enables the formation of Wh-questions. On a broad level, this study suggests that understanding question formation in more detail in first language acquisition may contribute insight into the theoretical enterprise of developing a theory of linguistic cognition. Since it makes use of naturally occurring language data, this paper additionally provides a basis for comparative study of such areas as bilingual development and impaired language development.

Hansun Zhang Waring's conversation-analytically informed study examines the discursive practices used for accomplishing disagreement and critique by participants in a graduate-level seminar discussion. Waring analyzes ways that novice seminar participants are able to both acknowledge another participant's viewpoint and simultaneously build a critique of that viewpoint within a single turn-at-talk within the institutional context of seminar discussions. This study adds to a small but rich tradition of analysis of discourse in academic meetings, and its findings

furthermore provide a valuable source for language educators concerned with the teaching of appropriate discourse practices to nonnative speakers engaged, or preparing to engage, in discussion within such academic settings.

We are especially pleased to offer our readers a special focus on language assessment in this issue. The later part of the issue includes a study examining the constructs of coherence in written discourse as well as interviews with two leaders in the areas of language testing and assessment research. Jungok Bae's study of coherence and cohesion addresses the practical but thorny problem of assessing the writing of elementary school students in a two-way bilingual immersion program, as well as those in English-only classes, through the use of a picture-based narrative writing task. Her study analyzes actual writing samples of such students in order arrive at quantifiable constructs for measuring cohesion and coherence, as well as for understanding their overall relationships to grammar and content. By using both monolingual and bilingual speakers in the study, Bae is able to shed some light on how the relationships among these elements of written text may be understood to be a part of an emerging ability in narrative writing across children with different linguistic backgrounds. This study has implications not only for the assessment of the writing of students in two-way immersion programs, but also for understanding the ways in which tests conceive of and measure writing ability more generally.

We are also fortunate to include in this issue interviews with two important figures in the field of language assessment, Charles Alderson and Dorry Kenyon. In these interviews, Viphavee Vongpumivitch and Nathan Carr explore current and emerging issues in the field of assessment as well as practical problems faced by language test developers within the constraints of real-world test projects. Among the highlights of the interview with Charles Alderson are a discussion of his seminal work on washback theory as well as his experiences with and views on the challenges and advantages of computer-based and web-based testing. Dorry Kenyon discusses a range of test development projects, including several related to the ACTFL scale of oral proficiency in foreign languages. He offers valuable insights into the issues and challenges faced in the development and validation of second and foreign language tests with the resources and constraints presented by various institutions seeking tests to use for particular purposes.

Finally, we would like to add that this issue of *ial* marks our third and final issue as co-editors. In addition to the range of new and developing research we have become acquainted with over the past two years, we have learned first-hand that the publication of a journal is a truly collaborative effort. The publication of this journal depends not only on the research of our many contributors, but also on the effort of many reviewers who contribute their time and judgment to the development of manuscripts, as well as the journal's staff, who put countless hours into keeping the organization running. Without the efforts of our graduate student staff, who volunteer their valuable time, *ial* could not remain in production issue after issue and year after year. We also want to welcome the new editors, Debra Fried-

man and Emmy Goldknopf, who have worked on the journal in various capacities over the past years. The next issue, Volume 12.2, will be their first at the helm as editors. We wish them the best of luck in carrying forward the long tradition of this student-run journal and look forward to the issues of *ial* yet to come.

June 2001

David Olsher
Leah Wingard

Wh- Movement in Child Catalan

John Grinstead

University of Northern Iowa

This study examines the production of wh- questions in the speech of four monolingual child speakers of Catalan who were recorded longitudinally as part of the study carried out by Serrà & Solé, obtained from the CHILDES Data Base (MacWhinney & Snow, 1985). In this data all wh- questions produced appeared to be adult-like, in contrast with the non-adult-like production of this construction in child English, Swedish, Dutch and German. However, there is an initial period in which no wh- questions at all are produced, in spite of the fact that other aspects of syntax, such as negation, clitic-placement and complementation seem adult-like in this same period. During this "no wh- question" period, there is a concomitant absence of verbal tense morphology, with the exceptions of present and irrealis forms (imperatives, root infinitives, root gerunds and root participles). Interestingly, the onset of wh- questions appears to correlate with the onset of a much wider variety of tense morphology. Given this observation and Rizzi's (1991) hypothesis that wh- questions and tense morphology are crucially linked in adult syntax, I propose that the early absence of wh- questions is a consequence of the early underspecification of tense.

What can child language development tell us about the principles that govern the domain of the mind dedicated to language? In answering this question, I will take the goal of formal linguistic theory to be the discovery of principles of mind. In determining what these principles are, all forms of relevant data should be brought to bear. For example, modules of grammar such as syntax and semantics can be selectively impaired in cases of aphasia (cf. Grodzinsky, 1990), suggesting the validity of a linguistic theory which divides syntax and semantics into distinct domains. In this way, an independently substantiated theoretical division between syntax and semantics is confirmed by neurobiological evidence. Similarly, in spite of all of the logically possible permutations of English auxiliaries, Stromswold (1990) presents evidence which suggests that children do not raise the lower auxiliary when forming questions, as in (2):

(1) You could have picked up the banana.

(2) *Have you could *t* picked up the banana?



(3) Could you *t* have picked up the banana?



This fact provides evidence for a principle of Universal Grammar (UG) known as Relativized Minimality (Rizzi, 1990) which governs how syntactic elements can move. Roughly, the principle says that things of the same type (verbal heads in this case) cannot move over one another when moving to a higher clausal position, as occurs, by hypothesis, in (2). Thus, study of child language data confirmed an independently established principle of Universal Grammar, Relativized Minimality. Consequently, a first question to ask about how questions are formed during the development of child Catalan is: Are there phenomena in child Catalan which might either confirm or falsify current conceptions of how questions are formed in adult Catalan? We will return to this question below.

Thus, the explanatory goal of this study is to contribute insight to the theoretical enterprise of developing a theory of linguistic cognition, as in generative grammar. Another important goal, however, is simply an accurate account of the facts of the development of child Catalan. This second goal is important because it makes basic research in child language relevant to other fields of linguistics, some of which have traditionally been called applied linguistic fields. For instance, without some clear idea of what the facts of monolingual child Catalan are, it is impossible to make judgements about the development of *bilingual* Catalan child language. That is, without normal, baseline data on child Catalan, it is difficult to judge whether bilingual Catalan children are developing Catalan on a normal linguistic schedule or not. Such monolingual data has been used in the development of the Independent Development Hypothesis (cf. Bergman, 1976, Paradis & Genesee, 1996). Similarly, in the field of communicative disorders, it is impossible to determine whether child Catalan syntax is developing on a normal developmental schedule without knowing what that schedule is in normally developing children, as in the four children studied here (cf. Anderson, 1999). In this way, careful descriptive studies of normal monolingual grammatical development are, for two fields of applied linguistics, crucial components of their work, without which their practitioners cannot hope to have a clear picture of developmentally impaired or bilingual development. The principle area of syntax to be addressed in this study is the development of question formation in child Catalan. Let us now turn to a description of the phenomena of non-adult like questions in child language.

It has been noted that children have difficulty producing adult-like questions in the early stages of their grammatical development. Child English speakers have been shown in various studies to either drop or fail to invert auxiliaries (Davis, 1987; Klima & Bellugi-Klima, 1966; Stromswold, 1990).

English

(4) What he can ride in? (Klima & Bellugi-Klima, Period 1)

(5) Where Ann pencil? (Klima & Bellugi-Klima, Period 3)

Child speakers of German, Dutch, and Swedish have also been reported to produce errors in question formation, although of a slightly different nature. These

children include the verb and place it correctly, but appear to leave out the wh-word itself (Swedish examples from Santelmann, 1995; Dutch from van Kampen, 1997; and German examples from Penner, 1994).

Swedish

- (6) *är det?* (Tor 19, 2;8)
is that? (Missing *vad*, 'what')
- (7) *kan den inte komma in?* (Tor 25, 2;11)
can it not come in? (Missing *varför*, 'why')

Dutch

- (8) *zeg je nou?* (Sarah, 2;1.19)
say you then? (Missing *wat*, 'what')
- (9) *heet zij nou?* (Laura, 3;5.30)
calls she then? (Missing *hoe*, 'how')

German

- (10) *isch das?* (S, 2;0)
is this? (Missing *wo*, 'where'/was, 'what')

The aim of the present article is to first determine whether or not errors of this kind occur in child Catalan and, if so, to explain them. To begin with, one must understand that to form an object wh- question in Catalan, the verb is placed before the subject and the wh- element is placed at the beginning of the sentence, as in (12). Consequently, if child Catalan speakers made errors similar to those made in child English, we might expect them to produce sentences like (13), with an uninverted verb, or (14), with a missing auxiliary. If they made errors similar to those made in child Swedish, Dutch and German, we might expect them to make errors such as that exemplified in (15), with a missing wh- element.

Catalan

- (11) *En Joan ha menjat la poma.* (Adult-like Declarative)
Joan has eaten the apple.
- (12) *Què ha menjat en Joan?* (Adult-like Question)
What has eaten Joan?
- (13) *@Què en Joan ha menjat?* (Question with no Subject-Verb
Inversion - Unattested)¹
What Joan has eaten?
- (14) *@Què en Joan menjat?* (Question with a Dropped
Auxiliary - Unattested)
What Joan eaten?

(15) @*En Joan ha menjat?*
Joan has eaten?

(Question with a Dropped Wh-
Word - Unattested)

However, we will see below that child speakers of Catalan do not make errors of either the kind found in child English or of the kind found in child Swedish, Dutch and German. Similarly, Guasti (1996) reports that child speakers of Italian, a related Southern Romance language, appear to produce only adult-like, obligatorily inverted wh- questions from very early. On the basis of these facts, Southern Romance child languages would appear to constitute an exception to the trend found in other child languages with respect to wh- question errors in that there are none. Children appear to simply behave as adults do from the very beginning.

Nonetheless, in this study I will argue that this “exceptionality” of child Southern Romance languages is only apparent and that in fact child speakers of Catalan (and perhaps Italian) do have a grammatical deficit vis-à-vis question formation. I will argue that this deficit prevents them from producing any wh- questions whatsoever in the early stage, in contrast to the child speakers of the Germanic languages mentioned, who are able to produce wh- questions but produce them incorrectly. Further, I will suggest that a particular part of clause structure, the Tense Phrase, is unspecified in early child Catalan and that this underspecification is responsible for the early absence of questions. By *underspecification* I mean that the part of the clause to which verbs move to attach to a tense morpheme is present, but has no syntactic, semantic, or phonetic content.

In the second section, I will establish that there is, in fact, a deficit in early wh- question production. This is important because research on the development of questions in children has suggested that wh- question production is adult-like. In the following section, I will explore two possible explanations for the early absence of wh- questions. First, I will address the possibility that this deficit is due to general syntactic immaturity in the speech of the Catalan-speaking children. Then, I investigate the possibility that the absence in the child grammar of the part of clause structure to which wh- elements are hypothesized to move in the adult grammar, CP (the complementizer phrase), is the cause of the deficit. Next, I suggest that an adult theory of wh- question formation (Rizzi, 1991) provides a promising line of inquiry into the cause of the deficit. I then briefly discuss the particular grammatical model I assume (Chomsky, 1995) and the way in which Rizzi’s theory may be interpreted in this model. Finally, I suggest that by assuming a crucial dependence between the development of the Tense Phrase in child Catalan clause structure and the development of wh- questions, the early absence of wh- questions in child Catalan (and possibly child Italian) can be explained.

A WH- QUESTION DEFICIT IN CHILD CATALAN

Do child speakers of Catalan have a deficit in their production of wh- questions? Asking this same question for child Italian, Guasti (1996) reports that the child Italian production data is error-free with respect to wh- questions. Guasti looks specifically at whether child Italian speakers fail to invert subjects and verbs, as do child English speakers. She reports that three Italian-speaking children of the Calambrone corpus, whom she studied, never produced an uninverted, hence non-adult-like, question, as in (16).²

- (16) @*Che Gianni ha mangiato?*
What Gianni has eaten?

Of course because Italian is a null-subject language,³ one can only determine whether subject-verb inversion has failed to take place if an overt subject has been used. According to Guasti (1996), out of 171 spontaneous utterances produced by Martina (1;8 - 2;7), Diana (1;10 - 2;6), and Guglielmo (2;2 - 2;11), 67 had overt subjects. Of these overt subject questions, which are the ones capable of telling us whether or not inversion has occurred, only three included non-inverted word orders, and these were *perchè* (why) questions which are grammatical in the adult language without inversion. These results are summarized in Table 1.

Table 1: Wh- Questions, Wh- Questions with Overt Subjects, and Inverted Wh- Questions with Overt Subjects in Three Speakers of Child Italian from the Calambrone Corpus (Sumarized from Guasti, 1996)

	Total # of Wh- Questions	Total # of Wh- Questions with Overt Subjects	Total # of Inverted Wh-Questions with Overt Subjects
Diana (1;10 - 2;6) Guglielmo (2;2 - 2;11) Martina (1;8 - 2;7)	171	67	64

Thus, Italian-speaking children appear to form questions in an essentially adult-like way from the beginning.

If we analyze wh- questions in child Catalan using Guasti's methodology, we find a similar result. That is, if we cull all questions asked by the four monolingual Catalan speaking children from the Serrà and Solé corpus from the CHILDES data base (MacWhinney & Snow, 1985), and then count the number of wh- questions with overt subjects with and without inversion we get the results in Table 2.⁴

As illustrated in Table 2, out of 146 wh- questions that these four children asked, 37 had overt subjects.

Table 2: Wh- Questions, Wh- Questions with Overt Subjects, and Inverted Wh- Questions with Overt Subjects in Three Speakers of Child Catalan from the Serrà and Solé Corpus of the CHILDES Data Base (MacWhinney & Snow, 1985)

	Total # of Wh- Questions	Total # of Wh- Questions with Overt Subjects	Total # of Inverted Wh-Questions with Overt Subjects
Gisela (1;7 - 3;0) Guillem (1;0 - 3;1) Laura (1;7 - 3;3) Pep (1;0 - 3;0)	146	37	37

All of these 37 questions had either a left dislocated subject, as in (17), or an utterance-final subject, as in (18), both of which are well-formed in the adult language.

(17) *Papa on és?* (Guillem - 2;11.25)
papa where is
Where is Papa?

(18) *On està la groga?* (Gisela - 2;8.0)
where is the yellow
Where is the yellow one?

While looking at the data from Guasti's (1996) perspective tells us that the Italian and Catalan children do not make the same errors that child English speakers make, it may not tell us anything about errors that might be specific to these particular child languages. That is, we might expect that the errors that occur in child Southern Romance, if there are any, will be different from the child English errors because the errors we find in child German, Dutch, and Swedish are also unlike the child English errors.

So, to examine the Catalan children's development in greater detail, each verbal utterance in each file was coded for tense and illocutionary force (question, statement, command, etc.). Then, using text search utilities on a UNIX computer, the number of occurrences of each code was calculated for each file. Using this procedure to calculate the total number of verbal utterances per file and the total number of wh- questions per file, we get the results in Table 3.

What stands out in Table 3 is that there is a lengthy period in the data of all four children before they produce any wh- questions at all. In Table 3, there are two columns after each child's age. The first of these gives the number of verbal wh- questions for the particular file and the second gives the total number of verbal utterances produced in that file. The numbers in bold are considered to be the first non-formulaic wh- question which included a verb for each child. The wh- questions occurring before those in bold, in the data of Guillem and Laura, are

Table 3: Verbal Wh- and Total Verbal Utterances (VU) in Four Monolingual Catalan-Speaking Children

Gisela			Guillem			Laura			Pep		
	Wh-	VU		Wh-	VU		Wh-	VU		Wh-	VU
(1;7,14)	0	0	(1;0,0)	0	0	(1;7,20)	0	0	(1;0,27)	0	0
(1;8,3)	0	8	(1;1,23)	0	0	(1;9,7)	0	6	(1;1,28)	0	0
(1;8,24)	0	11	(1;1,29)	0	0	(1;10,22)	0	25	(1;3,23)	0	0
(1;9,0)	0	4	(1;4,18)	0	0	(1;11,12)	0	37	(1;4,24)	0	5
(1;10,7)	0	10	(1;4,26)	0	0	(2;5,5)	0	25	(1;5,29)	0	13
(1;11,11)	0	2	(1;5,29)	0	0	(2;2,13)	1	52	(1;6,23)	0	10
(2;1,23)	0	7	(1;6,26)	0	2	(2;4,11)	1	13	(1;8,0)	0	7
(2;2,6)	0	0	(1;7,15)	0	2	(2;5,8)	0	72	(1;8,30)	0	12
(2;4,25)	0	49	(1;7,22)	0	0	(2;6,25)	4	41	(1;10,6)	0	73
(2;6,23)	0	29	(1;8,0)	1	14	(2;7,20)	1	120	(1;11,6)	4	50
(2;8,0)	19	224	(1;9,12)	0	26	(2;8,30)	1	148	(2;0,0)	0	18
(2;9,16)	8	154	(1;9,24)	0	11	(2;11,17)	7	157	(2;1,1)	1	59
(2;11,0)	4	103	(1;11,13)	0	36	(3;0,2)	5	293	(2;2,3)	5	92
(3;0,29)	0	23	(2;0,12)	0	16	(3;3,21)	1	220	(2;3,10)	3	106
			(2;1,14)	0	27				(2;4,4)	5	95
			(2;2,11)	0	8				(2;5,4)	7	163
			(2;2,28)	2	19				(2;6,15)	1	19
			(2;3,12)	0	4				(2;7,8)	2	121
			(2;3,18)	0	26				(2;7,28)	0	12
			(2;4,24)	4	37				(2;9,10)	7	197
			(2;5,25)	4	30				(2;10,15)	4	99
			(2;5,29)	1	24				(2;11,10)	0	100
			(2;6,10)	10	35				(3;0,27)	1	114
			(2;7,9)	0	44						
			(2;7,25)	1	71						
			(2;9,8)	2	99						
			(2;10,3)	4	29						
			(2;11,5)	5	35						
			(2;11,21)	2	68						
			(2;11,25)	7	91						
			(3;0,0)	1	7						
			(3;1,18)	8	68						

possibly formulaic, lexicalized units, given in (19) and (20). The number of total utterances is provided to show that questions are not missing as a consequence of the child not producing any utterances at all.

- | | | |
|------|--------------------------------------|-------------------|
| (19) | <i>Què és això?</i>
what is that? | (Guillem - 1;8.0) |
| (20) | <i>Què és?</i>
what is (that)? | (Laura - 2;2.13) |

The reader will notice that Gisela produces 19 questions in the first file in which she produces any questions at all. This was not typical. In fact, nothing seems typical when it comes to the absolute number of wh- questions children produce. It is likely that non-grammatical considerations induced different numbers of questions in different sessions.

To be precise, only wh- words which occurred *with verbs* were counted as questions in Table 3. During the period before verbal wh- questions begin to be formed, there are, nonetheless, a wide array of wh- words which occur without verbs, as in Figure 1, (a) through (g). The importance of these utterances will become clear later.

Figure 1: Wh- Words in Child Catalan Before Verbal Questions Are Formed

- | | | |
|-----|--------------------------------------|-------------------------------|
| (a) | <i>Quin?</i>
which (one)? | (Gisela - 1;8.24) |
| (b) | <i>Este a on?</i>
this one where? | (Gisela, 1;8.24) ⁵ |
| (c) | <i>Què bé.</i>
how good. | (Guillem, 1;8.0) |
| (d) | <i>Per què no?</i>
why not? | (Laura, 1;11.12) |
| (e) | <i>Què?</i>
what? | (Laura, 1;11.12) |
| (f) | <i>On, on?</i>
where, where? | (Laura, 2;2.5) |
| (g) | <i>Quina!</i>
which one! | (Laura, 2;2.13) |

Similarly, in the speech of Rosa, one of the child Italian speakers of the Calambrone corpus which Guasti used, we also find a period of time before any wh- questions

are produced, as illustrated in Table 4. During this same period, wh- words without verbs sometimes occurred. We will return to the significance of these non-verbal question words below.

**Table 4: The Occurrence of Wh- Questions in the Speech of Rosa
(Calambrone Corpus, CHILDES Data Base,
MacWhinney & Snow, 1995)**

	Wh- Questions	Total Verbal Utterances
1;7.13	0	1
1;9.11	0	14
1;10.08	3	30
1;11.24	2	30
2;0.07	6	29
2;1.14	7	32
2;01.29	2	35
2;02.11	5	43
2;4.23	36	121
2;5.25	3	78
2;6.29	17	116
2;7.26	11	146
2;9.04	13	186
2;9.24	8	178
2;10.14	12	167
2;11.12	12	167
2;11.30	50	187
3;0.24	34	158
3;1.29	15	151
3;3.23	12	208

THE PERIOD PRECEDING VERBAL WH- PHRASES IS NOT PRE-SYNTACTIC

Why should it be that child speakers of Catalan do not produce any wh-questions early on? Could it be that their grammars are in general too unsophisticated or underdeveloped for them to be able to produce questions? It would seem

not. While many of the utterances in this pre-question period are simple imperatives and 3rd person singulars, others of them seem quite sophisticated and adult-like, as in (21) through (25).

- (21) *No n' hi ha.* (Gisela, 1;9.0)
 not CL (partitive) CL (locative) is.
 There isn't any of that.
- (22) *Vull beure.* (Guillem, 1;8.0)
 want (1st SG) to drink.
 I want to drink.
- (23) *Papa, vull provar-ho.* (Guillem, 1;8.0)
 papa, want (1st SG) to try CL (ACC SG MASC)
 Papa, I want to try it.
- (24) *Et dono això.* (Guillem - 1;8.0)
 CL (DAT 2nd SG) give (1st SG) that.
 I give you that.
- (25) *Dóna-me-la.* (Pep - 1;5.29)
 give (2nd SG IMP) CL (DAT 1st SG) CL (ACC SG FEM)
 Give it to me.

Notice that in Examples (21) through (25) partitive, accusative, dative, and locative clitics are used and that they are used in their correct verb-final position in infinitives and imperatives, as in (23) and (25), and in their correct verb-initial position with finite verbs, as in (21) and (24). Notice as well that negation seems adult-like, as in (21), as does nonfinite complementation, as in (22) and (23). In short, it does not appear to be the case that the period preceding verbal *wh*-questions can be characterized as pre-syntactic or as any kind of one-word stage. Why, then, are no questions used by any of these four children?

IS CP ABSENT?

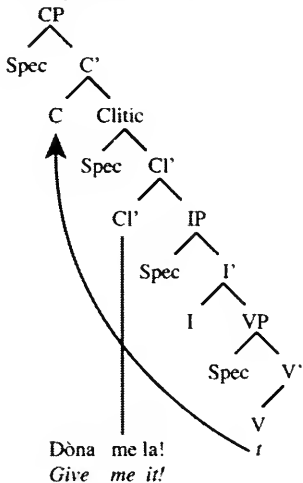
Some authors have suggested that question formation does not take place in the adult-like way in early child speech because the part of the clause to which *wh*-elements are hypothesized to move, the Complementizer Phrase (or CP), is not available, as in the work of Haverkort & Weissenborn (1991) for French and Meisel and Müller (1992) for German. Fortunately, in child Catalan we may test for the existence of CP, independently of *wh*-movement, by examining the use of imperatives. Following Rivero and Terzi (1995), I will assume that in adult Catalan, imperative verbs move to CP. Rivero and Terzi's assumption is based on the fact that imperative verbs precede clitics, as in (26), whereas finite verbs appear after clitics, as in (27).

- (26) *Dóna- me- la.*
 give (2nd SG IMP) CL (DAT 1st SG) CL (ACC SG FEM)
 Give it to me.(1)

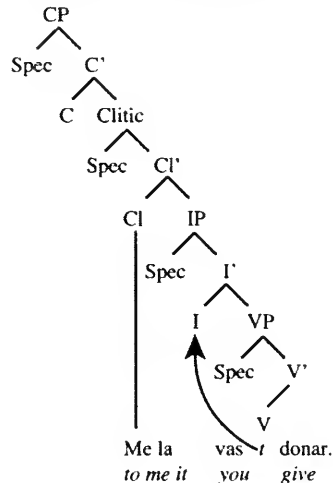
- (27) *Me la vas donar.*
 CL (1st SG DAT) CL (3rd SG FEM) aux (2nd SG PRET) give (INF)
 You gave it to me.

Thus, Rivero and Terzi (1995) assume that clitics occupy a stationary position in the clause structure and that imperatives move over them to C, as in (28), and that indicatives move only as high as the part of the clause hypothesized to hold inflectional material referred to as Infl (or just I), as in (29).

(28) Imperative Verb Movement to C



(29) Finite Verb Movement to I



From Rivero and Terzi's perspective (1995), which I adopt, any imperative a child produces is evidence for the existence of C. And, if imperatives can move to C, then it is not the absence of C from child Catalan clause structure which prevents questions from being formed. As we can see from Table 5, many imperatives occur in child Catalan before *wh*- questions begin to be used. To arrive at the number of imperatives out of the total number of verbal utterances produced, given in Table 5, I searched for imperatives in all of the children's files up to and including the file just preceding the first *wh*- questions. This number and percentage of imperatives suggests that the construction was extensively used, and, adopting Rivero and Terzi's assumption regarding the structure of imperatives, also suggests that C was present in the children's clause structures.

One might question, however, whether all of these imperative forms indeed involve movement to C. Noticing that most of these imperatives are 2nd person familiar imperatives, which are homophonous with third person singular indica-

Table 5: Number and Percentage of Imperatives Out of the Number of Total Verbal Utterances Produced by Each Child in an Early State of Child Catalan

	Number and Percentage of Imperatives Out of the Number of Total Verbal Utterances
Gisela (1;7 - 2;6)	31/120 (26%)
Guillem (1;0 - 2;2)	80/134 (60%)
Laura (1;7 - 2;2)	59/144 (41%)
Pep (1;0 - 1;10)	60/154 (39%)
Total	230/552 (42%)

tive forms, one might want to suggest that these are really “bare verb” forms in Catalan that are produced lower in clause structure which do not raise to C.⁶ However, if we adopt Rivero and Terzi’s assumption that clitics are positionally stable, then the occurrence of imperative verbs to the left of clitics constitutes evidence for the existence of C. In fact, all of the imperative forms that occur with clitics in this “pre-wh” stage occur to their left, as in examples (30) to (35), suggesting adult-like movement of imperative verbs to C.

Catalan

- (30) *Dóna-me- la.* (Pep, 1;5.29)
 give me (CL DAT) it (CL ACC FEM SG)
 Give me that.
- (31) *Busca-la.* (Pep, 1;6.23)
 seek it (CL ACC FEM SG)
 Look for it.
- (32) *Dóna 'm.* (Laura, 2;2.13)
 give me (CL DAT)
 Give me.
- (33) *Tornem- hi.* (Laura, 2;2.13)
 return (1st PL IMP) there (CL LOC)
 Let’s go back there.
- (34) *Tu dóna 'm iogurt.* (Guillem, 1;8.0)
 you give me (CL ACC) yogurt
 You give me yogurt.
- (35) *Ajuda 'm.* (Guillem, 1;9.12)
 help me (CL ACC)
 Help me.

Thus, it would appear that examples of imperative which occur with enclitics are neither “frozen forms” nor root infinitives and consequently constitute evidence of the existence of the C projection in child Catalan. Consequently, the inactivity of C is unlikely to be the reason for the lack of question formation in these child languages.

THE WH-/TENSE ASYMMETRY IN CHILD CATALAN

If neither general syntactic immaturity nor the unavailability of the C projection are responsible for the absence of questions early on in child Catalan, what is the cause of this apparent developmental delay? In this section I will suggest a possible explanation linking the delay in wh- question formation to a delay in grammatical tense marking.

The Wh- Criterion

A possible explanation for the inability of these children to form questions is that another aspect of child functional structure is not yet syntactically active, namely, Tense. Rizzi (1991) hypothesizes that Tense and wh- question formation are related in an important way. This hypothesis is based on the observation, following Haïk (1990), that there are languages in which an interrogative inflectional morpheme must always be present when questions are formed. Thus, Haïk notes that in Palauan, verbs in affirmative sentences carry an indicative mood morpheme (glossed as *R*), as in (36), while verbs in wh- questions carry an irrealis morpheme (glossed *IR*), as in (37).

Palauan

(36) Affirmative

Ng- kileld-ii a sub a Droteo
 R3s-heat- PF-3s soup Droteo
 Droteo heated the soup.

(37) Wh- Question

Ng-nerga a le-silseb-ii a se'el-il?
 cl-what IR-PF-burn-3s friend-3s
 What did his friend burn?

Rizzi (1991) proposes that this wh- morpheme, which is phonetically visible in Palauan and other languages, is in fact present in many (if not all) languages, though not phonetically visible. In Rizzi's (1991) formulation, this wh- morpheme is directly associated with the tense morpheme. This formulation is geared towards explaining why verbs (the auxiliary verb in the case of English) seem to raise above the subject in English and French in wh- questions. His idea is that the verb raises to the part of the clause where inflectional morphemes are added and at that point the verb acquires not only the tense morpheme, but also the invisible

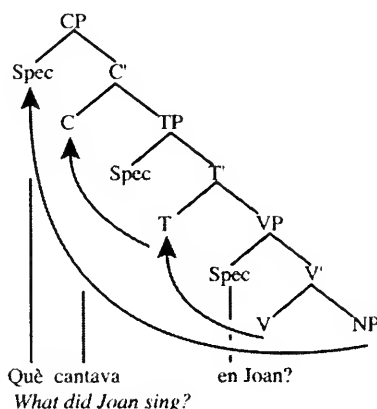
wh- morpheme. This wh- morpheme is bound by a condition that obliges it to be adjacent to the wh- question word (i.e., *what*, *where*, etc.). This is similar to the condition on direct objects in English that they must be adjacent to verbs.⁷ This adjacency condition requires the verb to move to the front of the sentence, where it can be adjacent to the wh- word. Rizzi calls this adjacency condition the *Wh-Criterion*, which is formalized in (38).

(38) **The Wh- Criterion** (Rizzi, 1991)

- a. A Wh- operator must be in a Spec-head configuration with a [+wh] X_0 .
- b. A [+wh] X_0 must be in a Spec-head configuration with a Wh operator.

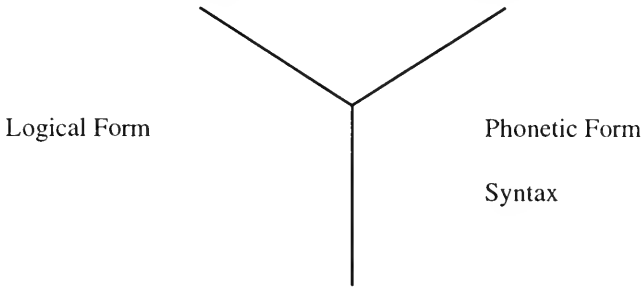
In (39), we see the verb move in two steps from the head of the Verb Phrase (V) to the head of the Tense Phrase (T), where it picks up both the tense morpheme and the wh- morpheme by Rizzi's (1991) hypothesis, and then moves to the head of the Complementizer Phrase (C). (The Catalan wh- phrase *què* moves to the specifier of the Complementizer Phrase).

(39)



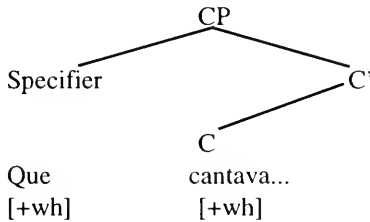
Minimalism

In *The Minimalist Program*, Chomsky (1995) attempts to develop an abstract theory to explain the fact that certain syntactic constituents appear to undergo movement within the clause. That is, the wh- pronoun *què* in object questions occurs at the beginning of the sentence, in spite of the fact that it is interpreted as if it were in object position, to the right of the verb. As a means of abstractly representing the motivation for this movement, Chomsky postulates a kind of feature that exists in the syntactic component of the grammar, which must be eliminated by the time that a syntactic derivation is given a phonetic form. Thus, in Figure 3, a syntactic derivation is generated which at some point branches off toward Phonetic Form, where it is pronounced. The idea is that these abstract features must be removed from the derivation before it passes to phonetic form or the derivation will "crash."⁸

Figure 2: The Computational Component in the *Minimalist Program*

According to Chomsky, these abstract features are eliminated by coming into a structural relationship with another element bearing the same feature. That structural relationship is known as the specifier-head relationship. In the case of *wh*-elements, and in Rizzi's (1991) formulation specifically, it means that a verb bearing an abstract *wh*-feature (carried by the tense morpheme) must move to the head of the Complementizer Phrase (C) and the *wh*-pronoun, also bearing an abstract *wh*-feature, must move to the specifier of the Complementizer Phrase (Spec, CP), as in (40).⁹

(40) The Specifier-Head Relationship



Implementing the *Wh*-Criterion in these terms, we see that both *wh*-pronouns and verbs with tense morphology must be present in order to produce a *wh*-question. Inferring from Chomsky's formulation of how these abstract features are eliminated, if either the tensed verb or the *wh*-pronoun is absent, then the abstract feature on whichever element is present will not be eliminated and will consequently cause the derivation to "crash." Thus, in Figure 2, a syntactic derivation is generated which at some point branches off toward Phonetic Form, where it is pronounced, while the rest of the derivation continues on to Logical Form, where the interpretation of the sentence is computed. Without one, the other cannot appear.

Wh- Words, Tense, and Feature Elimination

The examples in Figure 1, in Section 2, showed that during the early stage in which no syntactic questions are produced, children nevertheless produce wh- pronouns. Pep is the lone exception to this generalization. Hence there appears to be no lexical deficit with respect to the wh- words themselves that is preventing syntactic questions from being formed.

Thus, within the framework of assumptions just outlined, that leaves one other element in wh- questions which carries an uninterpretable feature, again, Tense. We can imagine that if Tense were not available to check the uninterpretable feature in these question words, then every derivation which included a question word with no tense should crash. Let us, then, examine the question of whether there is any overt morphological evidence for the existence of Tense as an active functional element in child Catalan. What we find in the way of verbal morphology in child Catalan before verbal questions are formed are second person singular imperatives, first and third person singular present indicatives, and a small number of root gerunds, infinitives, and participles, as in (41) through (44).

- (41) *Me 'n vaig.* (Pep, 1;4.24)
I (REFL) CL (PART) go (1st SG PRES)
I am going.
- (42) *Està aquí.* (Laura, 1;9.7)
is (3rd SG PRES) here.
It's here.
- (43) *Mira.* (Gisela, 1;10.7)
look (2nd SG FAM IMP)
Look.
- (44) *Dormir.* (Laura, 2;2.5)
to sleep.

As I argue in Grinstead (2000), these verb forms encode only present or irrealis temporal interpretation, which could be interpreted as the absence of tense. I will call this set of morphological tense markings *non-contrastive* to distinguish them from *contrastive* tense markings which I define as encoding speech time and event time as non-simultaneous. In Catalan, the contrastive forms include preterit, imperfect, simple future, periphrastic future, present perfect, past perfect, and the conditional. After an extended period during which only non-contrastive forms are used, contrastive tense forms begin to be used as well. I suggest that the onset of contrastive tense morphology is an indicator that syntactic and semantic tense specifications are then included in the child's syntactic structures.

Table 6: The Number of Files, Months, and Total Verbal Utterances Per Each Child's Early and Late Stage

	Number of Files	Number of Months	Total Verbal Utterances
Gisela I (1;7.14 - 2;2.6)	8	7	42
Gisela II (2;4.25 - 2;11.0)	5	7	559
Guillem I (1;5.29 - 1;9.12)	6	4	44
Guillem II (1;9.24 - 2;2.28)	6	5	117
Laura I (1;7.20 - 2;2.13)	6	7	145
Laura II (2;4.11 - 2;11.17)	6	7	551
Pep I (1;3.23 - 1;5.29)	4	5	18
Pep II (1;6.23 - 1;10.6)	5	5	102

Given the universality of some type of tense marking, I am inclined to believe that TP (the Tense Phrase) is a given part of the structure of the clause. The child's task is to specify the morphology that attaches to this category in their particular target language. Thus, I am assuming that TP comes unspecified for overt morphology, which must be learned, and that the abstract syntactic features I have referred to are not accessible to the child's grammar until the overt morphology has been learned.

To investigate the impact of the emergence of contrastive tense on the formation of questions we take the point at which the first contrastive tense morpheme is found in the speech of each child and search for questions both before and after that point for a roughly symmetrical number of files and months. For example, in the case of Laura in Table 5, the first contrastive tense morpheme is found in the seventh file, when she is 2;4.11. We then compare the preceding six files, which cover seven months, with the subsequent 8 files, which also cover 7 months. The number of files, months and total verbal utterances for both stages of each child, are given in Table 6. The point of dividing up the data in this way is to have roughly symmetrical periods of time intervals between recordings and roughly symmetrical amounts of recording time to compare. When we divide up the data in this way, we see that there is an early stage in which there are neither wh-questions nor contrastively tensed verbs. Then, sometime after the first contrastively tensed verb is produced, questions begin to be used. This is illustrated in Table 7.

Table 7: The Onset of Contrastive Tense Marking
and Verbal Wh- Question Formation

Gisela			Guillem			Laura			Pep		
	CT	Wh-		CT	Wh-		CT	Wh-		CT	Wh-
(1;7,14)	0	0	(1;5,29)	0	0	(1;7,20)	0	0	(1;1,28)	0	0
(1;8,3)	0	0	(1;6,26)	0	0	(1;9,7)	0	0	(1;3,23)	0	0
(1;8,24)	0	0	(1;7,15)	0	0	(1;10,22)	0	0	(1;4,24)	0	0
(1;9,0)	0	0	(1;7,22)	0	0	(1;11,12)	0	0	(1;5,29)	0	0
(1;10,7)	0	0	(1;8,0)	0	1	(2;2,5)	0	0	(1;6,23)	2	0
(1;11,11)	0	0	(1;9,12)	0	0	(2;2,13)	0	1	(1;8,0)	0	0
(2;1,23)	0	0	(1;9,24)	1	0	(2;4,11)	1	1	(1;8,30)	0	0
(2;2,6)	0	0	(1;11,13)	2	0	(2;5,8)	1	0	(1;10,6)	10	0
(2;4,25)	3	0	(2;0,12)	0	0	(2;6,25)	3	4	(1;11,6)	11	4
(2;6,23)	0	0	(2;1,14)	0	0	(2;7,20)	10	1			
(2,8,0)	13	19	(2;2,11)	0	0	(2;8,30)	12	1			
(2,9,16)	27	8	(2;2,28)	2	2	(2;11,17)	18	7			
(2;11,0)	23	4									

Having divided up the children’s data into two stages this way (a Pre-tense and a Tensed stage), we can compare the proportions of questions out of total verbal utterances in the Pre-tense stage with the proportion of questions out of total verbal utterances in the Tensed stage. Using the chi square test given in Table 8 on the following page to make this comparison, we find that the stages are significantly different. More specifically, this means that if the same ratio of questions to total verbal utterances existed in the Pre-tense stage as exists in the Tensed stage then we would expect there to be 8 or 9 questions in the Pre-tense stage, counter to fact. Thus, the stages appear to be qualitatively different vis-à-vis question formation. A possible conclusion to draw from these facts is that the use of contrastive tense morphology is necessary for questions to be formed. This possibility is made more plausible by the fact that the link between Tense and wh-questions has been proposed on different grounds, as in Rizzi (1991).

Table 8: The Ratio of Questions to Total Verbal Utterances in Two
Stages of Four Catalan-Speaking Children
Compared ($\chi^2 = 7.56$, $p < 0.01$)

	Questions	Total Verbal Utterances
Pre-tense	0	249
Tensed	47 (3.5%)	1333

SUMMARY

To summarize, we have seen that while wh- questions in child Catalan appear adult-like from the point at which they begin to be used, there is a lengthy period preceding their emergence during which no questions are formed. It seems unlikely that this lack of wh- questions is due to the absence or inactivity of C, given the imperative and clitic evidence presented. Furthermore, this deficit does not appear to be due to general syntactic immaturity because syntactically sophisticated constructions are used.

I have suggested that this complete absence of wh- questions can be explained by adopting Rizzi's (1991) hypothesized link between wh- questions and tense morphology. My implementation of Rizzi's hypothesized link makes use of mutual feature checking, as proposed in the *Minimalist Program* (Chomsky, 1995). Concretely, because children lack tense morphology early on, their clauses lack the ability to host one of the two wh- features necessary for mutual elimination of these features to take place. Thus, the children are left with wh- words, which carry uninterpretable features and lack the tense morphology which would carry the other wh- feature necessary to eliminate the feature of wh- words. As a result, all wh- question derivations that children attempt crash, and consequently no wh- questions are produced.

BACK TO ITALIAN

If the picture I have presented for child Catalan is correct, then a similar account of child Italian should be possible. The account developed here predicts the following:

- There should be an early period during which no verbal questions are asked.
- During this period, wh- words should, nevertheless, be available in children's vocabularies.
- The production of verbal wh- questions should not precede the use of contrastively tensed verbs.

To test these predictions, the files of Rosa, one of the children from the Calambrone corpus from the CHILDES data base (MacWhinney and Snow, 1985) were coded by a fellow graduate student, Stefano Vegnaduzzo, a native speaker of Italian, for Tense and wh- questions. After tallying the codes for Tense in Italian, we find a situation which is essentially identical to child Catalan. In the early files, present indicative and imperatives are the only verb forms used. In the first file, there are a number of wh- words used without verbs, as illustrated in (45).

(45) **Wh- Words in Rosa (1;7.13)**a. *chi!*

who!

b. *che?*

what?

As we can see in Table 9, these wh- words occurred in Rosa's vocabulary before the first use of a question or of a contrastively tensed verb, which both occurred in the third file.

Table 9: Contrastively Tensed Verbs and Verbal Wh- Questions in the Speech of Rosa (Calambrone Corpus, CHILDES)

	Passato Prossimo & Imperfect	Wh- Questions	Total Verbal Utterances
1;7.13	0	0	1
1;9.11	0	0	14
1;10.08	1	3	30
1;11.24	2	2	30
2;0.07	0	6	29
2;1.14	1	7	32
2;01.29	2	2	35
2;02.11	1	5	43
2;4.09	1	52	99
2;4.23	0	36	121
2;5.25	4	3	78
2;6.29	6	17	116
2;7.26	3	11	146
2;9.04	13	13	186
2;9.24	5	8	178
2;10.14	8	12	167
2;11.12	8	12	167
2;11.30	25	50	187
3;0.24	26	34	158
3;1.29	6	15	151
3;3.23	22	12	208

In the third file, the *passato prossimo* past tense begins to be used as *do* wh-questions. Strikingly, they not only begin to be used in the same file, but actually co-occur in the same utterance, given in (46).

- (46) *Detto baba detto cosa ha detto?*
 said papa said what has said
 Papa said, said, what did he say?

While further study of both Catalan and Italian is necessary to confirm or refute this hypothesis, the Italian child data nevertheless suggests that Italian-speaking children, like Catalan-speaking children, pass through a period during which no verbal question formation is possible. When contrastive Tense enters their grammar, elimination of the wh- feature carried by wh- words becomes possible. This in turn allows verbal wh- question derivations to be carried out, because the impediment which caused them to crash before this point can be removed.

CONCLUSION

We now have a theory of development which suggests that children need to acquire contrastive tense features before being able to check certain of the features associated with tense. I assume throughout that the features that cause wh- movement are available as part of UG. The child then has to learn the morphology associated with Tense in order to make available an attachment site for the wh-feature.

What this means for adult linguistic theory is that the connection between syntactic tense and wh- question formation proposed by Rizzi (1991) has been confirmed in child language development. More generally, the suggestion that there is an important connection between the syntactic location of tense in the clause and the Complementizer Phrase, first attributed to den Besten (1983), has similarly been confirmed. The descriptive contribution of the study is that in child Catalan, there is a period before questions are formed when other syntactic processes nevertheless seem to be active and that imperatives seem to be used before tensed verbs and questions are used. Imperatives seem to emerge developmentally with the group of other root verb forms which do not represent temporal distinctions other than with event time and speech time as simultaneous, such as gerunds, participles, and possibly some default forms which appear to be 3rd person singular present tense. My hope in this study has been to show that there is pattern in these data and that the pattern is explicable in terms of adult linguistic theory. If the reader is still skeptical of generative linguistic theory my hope is that the description of the data may prove useful nonetheless.

APPENDIX

MORPHEMIC GLOSSES

*-Ungrammatical
t-Trace
 @-Unattested
 CL-Clitic
 1st, 2nd, 3rd-Person
 SG-Singular
 PL-Plural
 FEM-Feminine
 MASC-Masculine
 DAT-Dative
 ACC-Accusative
 LOC-Locative
 PART-Partitive
 REFL-Reflexive

NOTES

¹ The symbol @ denotes an unattested utterance.

² The Calambrone corpora were collected naturalistically with regular visits to the homes of monolingual native speaking children of Italian (for more detail, see Cipriani, et al., 1989).

³ A *null subject language* is one in which the identity of the subject can be recovered either from grammatical information, such as the verb ending (e.g. only *io*, 'I', can be the subject of the verb *parlo*, by virtue of the verb's ending) or by discourse information. Crucially, an overt subject such as *I*, *John* or *the defendants* is not always necessary in such a language, as it is in English.

⁴ The Serrà and Solé study was conducted naturalistically by making monthly visits to the monolingual Catalan speaking children's homes and videotaping them. The transcripts were then donated to the CHILDES Data Base (MacWhinney and Snow, 1985). The Catalan files were downloaded from the CHILDES database and then searched for questions. The results are compiled in Table 2.

⁵ *Este* is a borrowing from Spanish.

⁶ In fact, I find this speculation appealing for explaining the root non-finite puzzle in child Catalan (Grinstead, 1998).

⁷ cf. Stowell (1981). Notice that the adverb *suddenly* can occur anywhere in the following sentences, except in between the verb and its object. Evidence like this suggested the Adjacency Condition, proposed by Stowell.

- (i) Suddenly, John hugged Grace.
- (ii) John suddenly hugged Grace.
- (iii) *John hugged suddenly Grace.
- (iv) John hugged Grace suddenly.

⁸ *Crashing* in this case would mean that the computational system had produced a derivation which included elements that the Phonetic Form component could not interpret, and consequently the resulting derivation would not be well-formed.

⁹ The point of developing a precise formal model of the structural properties of wh- question formation is to allow linguists to abstractly represent properties of grammar (and consequently of the mind) which hopefully interact with other grammatical patterns, allowing us to develop a principled theory of the mind in a manner analogous to the successful theories of natural sciences.

REFERENCES

- Bergman, C. (1976). Interference vs. independent development in infant bilingualism. In G. D. Keller & S. Viera & R. V. Teschner (Eds.), *Bilingualism in the bicentennial and beyond* (pp. 86-95). New York: Bilingual Press/Editorial Bilingüe.
- Chomsky, N. (1995). *The minimalist program*. Cambridge, MA: MIT Press.
- Cipriani, P., Pfanner, P., Chilosi, A., Cittadoni, L., Ciuti, A., Maccari, A., Pantano, N., Pfanner, L., Poli, P., Sarno, S., Bottari, P., Cappelli, G., Colombo, C., & Veneziano, E. (1989). *Protocolli diagnostici e terapeutici nello sviluppo e nella patologia del linguaggio* (1/84, Italian Ministry of Health): Stella Maris Foundation.
- Davis, H. (1987). *The acquisition of the English auxiliary system and its relation to linguistic theory*. Unpublished doctoral dissertation, University of British Columbia, Vancouver.
- den Besten, H. (1989). *Studies in West Germanic syntax*. Amsterdam: Rodopi.
- Grinstead, J. (1998). *Subjects, sentential negation and imperatives in child Spanish and Catalan*. Unpublished doctoral dissertation, UCLA, Los Angeles.
- Grinstead, J. (2000). Tense, number and nominative case in child Catalan and Spanish. *Journal of Child Language* 27(1), 119-155. Unpublished doctoral dissertation, UCLA, Los Angeles.
- Grodzinsky, Y. (1990). *Theoretical perspectives on language deficits*. Cambridge, MA: MIT Press.
- Guasti, M. T. (1996). The acquisition of Italian interrogatives. In H. Clahsen (Ed.), *Generative perspectives on language acquisition* (pp. 241-270). Amsterdam: John Benjamins.
- Haik, I. (1990). Anaphoric, pronominal and referential infl. *Natural Language and Linguistic Theory*, 8(3), 347-374.
- Havorkort, M., & Weissenborn, J. (1991). *Clitic and affix interaction in early child Romance*. Paper presented at the 16th Annual Boston University Conference on Language Acquisition, Boston.
- Klima, E. S., & Bellugi-Klima, U. (1966). Syntactic regularities in the speech of children. In J. Lyons & R. J. Wales (Eds.), *Psycholinguistics papers: The proceedings of the 1966 Edinburgh Conference* (pp. 183-208). Edinburgh: Edinburgh University Press.
- MacWhinney, B., & Snow, C. (1985). *The CHILDES Project*. Hillsdale, NJ: Lawrence Erlbaum.
- Meisel, J., & Müller, N. (1992). Finiteness and verb placement in early child grammars. In J. Meisel (Ed.), *The acquisition of verb placement* (pp. 109-138). Dordrecht: Kluwer.
- Paradis, J., & Genesee, F. (1996). Syntactic acquisition in bilingual children: Autonomous or interdependent? *Studies in Second Language Acquisition*, 18 (1-25).
- Penner, Z. (1994). Asking questions without CPs? On the acquisition of root wh-questions in Bernese Swiss German and Standard German. In T. Hoekstra & B. Schwartz (Eds.), *Language acquisition studies in generative grammar* (pp. 177-213). Amsterdam: John Benjamins.
- Rivero, M. L., & Terzi, A. (1995). Imperatives, V-movement and logical mood. *Journal of Linguistics*, 31, 301-332.
- Rizzi, L. (1990). *Relativized minimality*. Cambridge, MA: MIT Press.
- Rizzi, L. (1991). *Residual v2 and the wh criterion*. University of Geneva.

Santelmann, L. (1995). *The acquisition of verb second grammar in child Swedish*.

Unpublished doctoral dissertation, Cornell University, Ithaca, NY.

Stowell, T. (1981). *The origins of phrase structure*. Unpublished doctoral dissertation, MIT, Cambridge, MA.

Stromswold, K. (1990). *Learnability and the acquisition of auxiliaries*. Unpublished doctoral dissertation, MIT, Cambridge, MA.

van Kampen, J. (1997). *First steps in wh-movement*. Wageningen, Netherlands: Ponsen & Looijen.

John Grinstead has B.A. in Linguistics and Spanish, an M.A. in TESL and a Ph.D. in Applied Linguistics, all from UCLA. He works on the developmental syntax and semantics of child Spanish and Catalan, the development of numerical cognition and language disorders in monolingual Spanish and bilingual Spanish-English-speaking children. He is currently an assistant professor in the Department of Modern Languages at the University of Northern Iowa.

Balancing the Competing Interests in Seminar Discussion: Peer Referencing and Asserting Vulnerability

Hansun Zhang Waring
Mercy College

As Jacoby and McNamara (1999) have convincingly demonstrated, English for Specific Purposes (ESP) assessment tools with primarily a linguistic focus can fail to locate the competence actually needed in real-world professional settings. In a similar vein, English for Academic Purposes (EAP) pedagogical activities rooted in an unsituated notion of academic English can also be inadequate or misleading. Through a sequential analysis of actual interactions, this study describes the real-world discourse activities performed by competent native and nonnative speakers to handle complex academic tasks. Using data from a graduate seminar, I detail two interactional resources ("peer referencing" and "asserting vulnerability") exercised by the seminar participants in the doing of disagreement and critique. I show that these resources are invoked to accomplish the double-duty of acknowledging another's viewpoint while performing a potentially disagreeing action, to make an otherwise independently advanced critique into a co-constructed one, or to back down from forcefully articulated positions. Finally, I hypothesize that the particular use of peer referencing and asserting vulnerability characterizes the members' transitional stage between undergraduate novicehood and doctoral level junior expertise.

The pedagogical emphasis of English for Academic Purposes (EAP) has traditionally been placed upon academic writing (e.g., Leki, 1992; Raimes, 1999) and "academic lectures, formal speaking, or pronunciation" (Ferris, 1998, p. 291). Despite the difficulty nonnative speakers experience in class discussion (e.g., Jones, 1999), relatively less attention has been directed towards socializing this population into the discourse of multi-party interaction. Existing materials that purport to emphasize oral communication in the academy (e.g., Hartmann & Blass, 2000; Hemmert & O'Connell, 1998; Johns & Johns, 1977; Lynch & Anderson, 1992; Price, 1977; Steer, 1995) tend to prescribe lists of isolated linguistic expressions or functions which are believed to constitute "academic English." Generally lacking in this pedagogical orientation is an interactionally based understanding of the target discourse practices. As Jacoby and McNamara (1999) have convincingly shown, English for Specific Purposes (ESP) assessment tools with primarily a linguistic focus can fail to locate the competence actually needed in real-world professional settings. In a similar vein, pedagogical activities rooted in an unsituated view of academic English can also be inadequate or misleading. This study is intended as a preliminary endeavor aimed at informing the EAP pedagogical agenda by describing, through a sequential analysis of actual interactions, the real-world discourse activities performed by competent native and nonnative speakers to handle complex academic tasks.¹

In particular, I intend to examine the talk in a graduate seminar whose participants include both MA students and beginning level doctoral students who are still completing their course work. In the remainder of this paper, I use the term "graduate seminar" to refer to one with such mixed participants. Existing discourse analytic research on multi-party interaction that involves graduate student participation ranges from dissertation defense (Grimshaw, 1989), academic colloquium (Tracy, 1997; Tracy & Baratz, 1993; Tracy & Carjurzaa, 1993), to informal meetings among university physicists on a research team (Gonzales, 1996; Jacoby, 1998; Jacoby & Gonzales, 1991; Ochs & Jacoby, 1997) as well as graduate seminars (Brzosko-Baratt & Johnson-Saylor, 2001; Prior, 1998; Viechnicki, 1997; Waring, *in press*). The participants in the above events, with the exception of those in graduate seminars, are more likely to be graduate students at a fairly advanced stage of study. If graduate school experience can be construed as a continuum that indicates increasing scholarly maturity, it is perhaps fair to say that the research carried out so far has mostly illuminated the nature of talk at the advanced end of the continuum. Discussion in a seminar with a mixture of both MA and beginning level doctoral students would then contain the discourse practices that typically occur at the lower end of the continuum. In many ways, discussion in such a seminar is the common denominator of all other types of university talk in which graduate students participate. Unlike the other settings where graduate students' involvement is largely voluntary (e.g., colloquium), on a selective basis (e.g., research team), or a once-in-a-career occurrence (e.g., defense), a graduate seminar is the baseline experience of graduate school.

More importantly, a graduate seminar represents a crucial transitional stage between undergraduate novicehood and doctoral level junior expertise.² On the one hand, "collaborative discussion" has taken the place of "unidirectional informing" (Lakoff, 1990, p. 156). The students are no longer undergraduates entertained with information in a large lecture hall. For the graduate students, a considerable portion of the final grade often goes to class participation, where they routinely display their "fully-preparedness" through competent understandings of the readings. Learning is achieved at a higher level of independence, democracy, and collaboration, where interpretations of the reading materials are attempted, deliberated, clarified, contested, and fine-tuned. On the other hand, unlike other settings such as a doctoral seminar where the role of graduate students is more of independent, responsible, and contributing researchers, a reading seminar with mixed participants remains a guided learning event with clearly delineated goals articulated in a course syllabus. Unlike a doctoral seminar that primarily focuses on the participants' own original research being prepared for dissertation, conference presentation or publication, the central task in a graduate seminar involves the discussion of readings from books or journal articles. The professor selects the readings, steers the discussions, and evaluates the students' performance with a course grade.

As such, a graduate seminar is replete with complex interactional issues that reflect the members' transition from receptive undergraduates to thoughtful junior scholars. These issues tend to revolve around the doing of disagreement and critique. By proffering assessments through disagreeing or critiquing, one claims knowledge of that which is being assessed (Pomerantz, 1984). Graduate students in a reading seminar who are to both comprehend and critically respond to the works of published writers are, however, inevitably confronted with the somewhat paradoxical task of assessing that which they are still in the process of learning. Relatedly, their role as guided learners in a semi-structured speech exchange system where turns of talk are neither locally managed (like in an ordinary conversation) nor completely pre-allocated (e.g., traditional classroom) (Viechnicki, 1997, p. 105) makes it necessary for them to remain receptive, abiding participants on the one hand, and try to become more independent and thoughtful individuals on the other hand—through disagreeing or critiquing. To a certain extent, this paradox is perhaps also related to what Viechnicki (1997) calls "the inherent tension between" "self-aggrandizement and self-deprecation" (p. 111) that students face in a seminar. Similarly, Tracy (1997) points to the dilemma of looking "intellectually able without being seen as a show-off" among the discussants in the academic colloquia (p. 29). The purpose of this study is to examine two of the interactional devices deployed by the seminar participants to cope with the above dilemmas: peer referencing and asserting vulnerability. I will detail the composition, position, and action of these two devices, and discuss their properties in light of the unique context of a graduate seminar.

THE DATA SET

Data for this study consist of five weekly meetings of a nine member (professor included) graduate seminar titled "Trends in SLA: Second Language Reading and Literacy-Theory and Practice" at Teachers College, Columbia University in the fall of 1997. Each meeting lasted 1.5 hours long. Among the nine members, six were native speakers of English (Professor, Libby, Kelly, Ellen, Sam, and Jack³), and three spoke English as a second language (Ling from Taiwan, Kim from Korea, and Tamar from Israel). Libby, Ling, and Tamar were early doctoral students who were still doing their course work in the Applied Linguistics and TESOL programs, and the rest were MA students. Kelly was the only beginning level Master's student.

The seminar took place in a regular classroom. Before each session, the students moved the desks around to form a slightly rectangular table around which everyone could then sit and talk. In general, the professor opened the session with some overall comments regarding what was expected to be accomplished within the next 90 minutes. She also suggested the order of presentations and the approximate amount of time allotted to each presenter. Sometimes, the professor would begin a session with a guiding discussion question. For most sessions, two

assigned speakers presented two different articles that the entire group had read for a particular session. The speakers had been asked to summarize the articles and generate discussion questions. Discussion followed or interwove with the presentations. Each session ended with some concluding remarks by the professor.

The topic of discussion for each session was pre-determined in the course syllabus, varying from models of L2 reading theory, L1 versus L2 reading, L1 skill transfer, text-driven components of reading in L2, to knowledge-driven components of reading and L2 reading pedagogy. One of the sessions was also devoted to conducting an experiment in understanding the notion of "main ideas." According to the professor, the chosen readings were either seminal articles on key issues in the field or important materials related to the discussed topics. In short, the professor selected the readings which she believed the students ought to be familiar with in order to become knowledgeable about the field of second language literacy.

Five meetings were audiotaped, transcribed, and analyzed using conversation analysis (henceforth CA) (see Appendix for transcription notations). The first three meetings were also videotaped. (Further videotaping could not be arranged due to scheduling conflicts.) The video data were sometimes brought in to clarify the analysis, but were mostly used as backups for checking accuracy or identifying speakers. This research is based on the assumption that "one instance is sufficient to attract attention and analysis" because it is "an event whose features and structure can be examined to discover how it is organized" (Psathas, 1995, p. 50).

PEER REFERENCING

I use the term *peer referencing* to label a specific class of items found in the data (7 instances) where the expression "as/like you said" is placed either prefatory to or parenthetically inside the allegedly reported talk. This practice has been referred to in the communication literature as "naming," "referencing back" (Barnes & Todd, 1995) or "idea crediting" (Tracy, 1997). Peer referencing might also have something in common with "reported speech" (e.g., Holt, 1996; 2000; Li, 1986), "formulation" (Heritage, 1985; Heritage & Watson, 1979; Hutchby, 1996) or "reformulations" (Gonzales, 1996). Research on these phenomena has shown that reproducing what someone else has said before can accomplish a variety of actions beyond simple reproduction, such as conveying the current speaker's attitude towards the reported talk (e.g., Li, 1986; Gonzales, 1996; Holt, 1986; 2000; Myers, 1999). The actions performed by formulating or reformulating, for example, include demonstrating comprehension (Heritage & Watson, 1979), summarizing (Heritage, 1985), eliciting further talk (Heritage, 1985), indicating agreement (Heritage, 1985), or constructing a pre-move to disagreement (Gonzales, 1996; Hutchby, 1996).

Most notably, peer referencing is distinguished from formulations, reformulations, or reported speech by its design. Unlike formulations (“you don’t think X”) in Heritage (1985) and reformulations (“you’re saying X”) in Gonzales (1996) or “you say X” in Hutchby (1996), in peer referencing, the allegedly reported talk X is either prefaced or parenthetically marked by an adverbial clause such as “as you said” or “like you said.” As an initial observation, a free standing “You say X” or “You’re saying X” is primarily heard as an attribution which is often disaffiliative in its orientation (e.g., Gonzales, 1996; Hutchby, 1996). By contrast, the word “as” or “like” in “as/like you said, X” suggests a sense of similarity, convergence, or affiliation between the speaker and the reported talk. In what follows, I attempt to go beyond this initial observation by illustrating two particular actions performed by peer referencing in its sequential contexts—what I call *collaborative critique* and *inclusive disagreement*. In collaborative critique, one turns an otherwise independently advanced critique into a co-constructed one (6 instances). In inclusive disagreement, one acknowledges another’s view while constructing a potentially disagreeing position (1 instance).

Collaborative Critique

In the seminar data, one use of peer referencing is to collaboratively construct a critique. In the following segment, Ellen begins her turn by alluding to the author’s questionable choice of data. Her critiquing orientation is captured in negative assessments such as “He doesn’t really deal with them separately” (line 5) or “there’s no (.) data” (line 9). By noting that the author has lumped children and adults into a common category and applied evidence from native speaking children to second language learners, Ellen questions the internal validity of the study. And by pointing to the lack of data, she questions its external validity. The heart of her criticism, however, is couched in peer referencing (line 7):

[01] [1118.22]

- 1 Libby: Okay well my first question was the um what kind of assumptions does
 2 Krashen make about learners. For example children versus adults and native
 3 versus nonnative. Uhm °what do you think.°
 4 (3.0)
 5 Ellen: °He doesn’t really (.) deal with them separately. He treats them° (.) as learners
 6 and he uses the evidence of native (.) children (.) that’re I mean,
 7 —> as you said before supposed to apply to second language but
 8 (2.0)
 9 there’s no (.) ()data.
 10 (0.6)
 11 Kelly: And by combining them? One thing I- I was- con- confused about is, how does
 12 a nonnative speaker use (.) free (.) reading. (.) from (.) the beginning (.) where
 13 Kelly: they haven’t acquired any vocabulary to use as [()
 14 Tamar: [You jst- quoted this article.

Indeed, earlier in the discussion, Libby voiced her objection to using data from native speaking children to draw conclusions for second language reading. Ellen acknowledges that in peer referencing. By saying "as you said before," she presents Libby as the co-author and co-principal (Goffman, 1981) of the critique, and thus reminds everyone that she did not initiate the criticism of a widely popular scholar in the field. At the same time, she also uses what Libby has said before as a piece of substantiation for her own assertion of understanding. She thus strengthens her current response by indicating that the critique is not isolated and idiosyncratic, but rather a shared sentiment.

The role of peer referencing in forging a co-constructed, thereby strengthened, critique is captured with even greater clarity in the following example. At the beginning of the session, Kelly (the presenter) briefly mentions that the third component of the reading model under discussion is not something the author has addressed at any length. Later on, Sam refers back to Kelly's comment on the "third component" and reformulates it as a critique. Below is what Kelly said first:

[02] [1014.1]

- 1 Kelly: And the third component °>which he doesn't use very much in the study except
2 in the conclusion<°would be:: .hhh the outside factor which doesn't have to
3 do with the schemata o:r the text. They are the (.) ∞you know∞ (.) purpose the
4 studen- why the student took reading the purpose of reading the motivation
5 factors an- (.) emotional factors, anything else that plays a role.

Later on, as shown in the segment below, Sam explicitly states that the author failed to address in detail the third component of reading: "affect." The expression "like you said before" is again placed between the contextualizing background "the third component, which is more of the (.) affective" and the specific point of critique:

[03] [1014.7]

- 1 Prof: [lines limited] Does one help sometimes? when the other one fails? Okay?
2 Which is it that's (.) that's (.) you know (.) most useful in (.) in helping (.)
3 comprehension. Okay? Because ultimately, in all of these they all had a test at
4 the end.
5 Kelly: Right. [Right.]
6 Prof: [Right?]
7 (0.4)
8 Sam: I had a (.) question (.) but uhm (.) he also- the third component, which is more
9 —> of the: (.) affective. and js- like you said before, he seems to completely (.)
10 uhm leave that out. And at the end, he says that uh (.) he's talking about that
11 .hhh (.) motivational factors may lead advanced level readers to utilize skills
12 which other lower-level readers have so he seems to be (.) maybe there's (.) three
13 parts interplaying (by and large here)but this article only seems to be focusing
14 (.) on the interplay between these two but then he >kind of throws it in with the

- 15 others and says< oh yeah! it's these two that we (.) [kinda brought it up.]
 16 Kelly: [they could affect]
 17 everything.
 18 Sam: Yeah! you know

Although earlier on Kelly mentioned the fact that the author did not deal with the third component “very much” except in the conclusion (lines 1-2 in [02]), “completely leave that out” (lines 9-10 in [03]) was certainly not her wording. She provided a descriptive account. She also de-emphasized that account by producing it parenthetically in decreased volume and increased speed (lines 1-2 in [02]). Sam, on the other hand, calls attention to the same fact in a tone that ostensibly suggests a critiquing stance. By using expressions such as “leaving out” (line 10) something and “only” (line 13) focusing on two of the three components, he clearly alludes to some type of failure on the author’s part. In other words, Sam is the “principal” (Goffman, 1981) of, or the one whose position is represented in, the critique, not Kelly. Through peer referencing, he brings in a factual point mentioned by Kelly in passing, and remodels it in a fashion favorable to the critique he is attempting to construct.

Interestingly, Kelly does not seem to find the twist objectionable. She in fact proceeds to proffer a collaborative completion (line 16) of what Sam has begun. Considering the potential benefit of critiquing in indexing intellectual competence, Sam is arguably putting Kelly in an “indebted” position rather than acknowledging indebtedness himself. More importantly, however, the allusion that Kelly said this before also does Sam a favor by lending support to his argument under construction—one that Sam might not otherwise feel confident enough to advance single-handedly.

Compared to “as you were saying” which highlights the distance between the speaker and the proposition involved, “jst- like you said before” stresses that what is being proposed is not new or ungrounded. In light of the fact that Kelly’s original remarks were not critical in their orientation, it is even more plausible that peer referencing is used strategically by Sam to strengthen his own critique by turning an otherwise independent argument into a co-constructed one.

Inclusive Disagreement

I have also found one instance in which peer referencing is used not for *collaborative critique*, but for *inclusive disagreement*—a practice of acknowledging another’s view while constructing a potentially disagreeing position. The following segment is part of a larger discussion on a key finding in an assigned article. The article states that readers of limited linguistic proficiency can compensate for their language difficulties by using their background knowledge. Immediately prior to the segment was a sequence of talk initiated by Ellen who expressed concern about what seemed to be an overly tentative manner in which the author presented the aforementioned finding. The professor concluded that sequence by commending the author’s cautiousness and suggesting that the readers have the

freedom to interpret the finding in any way they deem appropriate. And that is exactly what Tamar proceeds to do at the beginning of the segment. She attempts to interpret the finding in practical terms, entertaining the possibility that the author might endorse a teaching strategy which emphasizes pre-reading background introduction:

- [04] [1014.10]
 1 Tamar: So-) (*adjusts her posture to look towards Kelly*) (if you look at it in practical
 2 terms? classroom? recommendation) Does that mean that when I have beginner's
 3 class, uh:m like this, a:h strategy would be: to: discuss the reading ahead of
 4 ti:me, and sa:y we'll a:hm (.) give a few ke:y (.) not WORDS, but key notions
 5 about the backgro:und, whatever, let's say I have a story about somebody >did a
 6 magic pray for rain whatever it says (.) sometimes there's too much rain,
 7 sometimes there's little rain, an' sometimes people think we can influence (.)
 8 uh..heaven whatever a discussion of it< and then going to the text? would that
 9 be a recommendation he would (.) uh:m if he hadn't been so cautious,
 10 [((laugh)) (w'd he)]
 11 Kelly: [I think that he-] (.) also almost like saying, if you expe:rience this (.)
 12 breakdown, that (.) this is the stra:tegy (.) to u:se, an:d (.) °(what worked).°
 13—> ((looks towards the professor who responds with a nod of approval)) Or
 14 >maybe with students like you were saying °of a certain level of proficiency°
 15 you can just use it as a< stra:tegy.
 16 Libby: .hhh °I think it's- oh I'm sorry.°=
 17 Kelly: =°Go ahead.°
 18 Libby: I think he's not saying u:se it, and ignore the others. He's saying u:se it in
 19 conjunction with the others. because there seems to be this- seasa:w, you know,
 20 of (.) the teachers continue to teach the lingui- it's the linguistic (.) element, and
 21 then the researchers >are saying don't do the top down do the top down.< so
 22 some people give up on the linguistic stuff, focus on the top down, and nobody
 23 knows how to read. so it has t- everything has to be done (.)
 24 [eh. (.) together in a wa:y,]=
 25 Kelly: [That's basically what he's saying.()]

As Tamar begins speaking, it is clear from the video data that she has selected Kelly the presenter—the “knower” of the session, as the primary, if not the sole, recipient of her talk through her gaze (Goodwin, 1981). Immediately after the turn initial “so,” she adjusts her posture to position herself so that she could face Kelly—the relative “knower” of the session. This body movement is unambiguously visible as Tamar has been sitting side by side with Kelly at the table facing the same direction. In addition, although Tamar's gaze travels around the table as her turn at talk unfolds, it repeatedly comes back to land on Kelly as her turn approaches the various focal points, such as “discuss the reading ahead of time” (line 3), “give a few key...notions” (line 4), and “a discussion of it” (line 8). Finally, Tamar ends her question “Would that be a recommendation...” (lines 8-9) by again looking directly at Kelly.

Tamar's multi-unit turn is initially set up by the "if-" clause which announces that she is about to discuss the author's findings with regard to "practical terms" or "classroom recommendation" (line 1). She begins with "Does that mean..." (line 2), asking whether discussing readings ahead of time would be a teaching strategy consistent with the author's stance. The possible completion point after the phrase "ahead of ti:me" towards the end of line 3 is the first turn transition opportunity passed up by Kelly (Sacks, Schegoff & Jefferson, 1974), after which Tamar continues with an expansion where she specifies what she meant by "discussing"—giving "key notions about the background" (lines 4-5). In other words, Tamar addresses the possibility that Kelly's withholding of response at the first possible completion point arises from her trouble understanding the question. As Tamar reaches a second possible completion point with the word "backgro:und" in line 5, Kelly again passes up her opportunity to speak, upon which Tamar utters "what-ever" to create a new transition-relevance place. When that fails to incur any uptake from Kelly, Tamar proceeds with yet another expansion (lines 5-8) where she uses "a magic pray for rain" as an example of the type of text to which a pre-reading discussion might be applied. As shown, up to "and then going to the text?" in line 8, Tamar has been focusing on clarifying her initial question as if to remove the road blocks that might have hindered Kelly from responding.

Having seemingly exhausted her resources of clarifying by this time with Kelly continuing to remain silent, Tamar goes on to remind her of the original question "would that be a recommendation..." (lines 8-9). It can be argued that the three subsequent items following the restated question (i.e., "he would," the micropause, the filler "uh:m") in line 9 constitute "monitor spaces" (Davidson, 1984, p. 117) where Tamar can assess the nature of Kelly's implicated response based on what happens or does not happen there. It seems that within these monitor spaces, Tamar has reanalyzed Kelly's withholding of uptake from her having trouble understanding the question to having difficulty agreeing with the recommendation. The fact that Tamar reconsiders her interpretation that the lack of uptake from Kelly is not surprising since Tamar's successive attempts at clarifying did not seem to have made a difference in shaping Kelly's response. The revised understanding (that Kelly had difficulty agreeing with the recommendation) is made evident in the ensuing qualifying clause "if he hadn't been so cautious" (line 9), which addresses Kelly's possible concern that such a recommendation would not be consistent with the author's findings, especially when those findings have been stated with great caution. Thus, by saying "if he hadn't been so cautious," Tamar in a sense relaxes the standard for admitting the "recommendation" and makes it more easily acceptable, thereby exhibiting an active orientation towards inviting agreement from Kelly.

It is conceivable that Tamar's earlier expansions were not intended as clarifications, but simply produced to create new transition-relevance places for Kelly to proffer the preferred response "yes" already built into the "yes/no" question "Does that mean..." (line 2).⁴ Tamar's question (in line 2) also invites agreement

in the sense that it is more an assessment than a question. Tamar does not simply say, "Would this have implication for classroom teaching?" Rather, she puts forward a candidate measure of application—an assessment whose preferred second assessment would be agreement (Pomerantz, 1984). In any case, regardless of whether the invitation for agreement was implied earlier on in Tamar's turn, it becomes clearly visible when she says "if he hadn't been so cautious."

What Kelly subsequently produces as she finally takes her turn is an elaborately designed two-part response suggesting that inducing schemata to compensate for linguistic limitations is first a reading strategy (lines 11-12), and then perhaps, as Tamar suggests, a teaching strategy used with students of a certain level of proficiency (lines 13-15). There are two points to be immediately noted about the first part of Kelly's response (lines 11-12). First, as she begins talking, the preferred/invited agreement is noticeably absent. Or, we could perhaps say it is further delayed or withheld (i.e., after the multiple transition-relevance places within Tamar's previous turn). From a sequential perspective, such absence or delay is disagreement-implicative (e.g., Pomerantz, 1984; Sacks, 1987). Second, complicating this disagreeing orientation is roughly a re-assertion of Kelly's own interpretation (lines 11-12) she presented earlier—a counter-interpretation of a sort. In other words, Kelly is not merely showing trouble granting Tamar's "recommendation" instant endorsement, her specific trouble has to do with the fact that Tamar's interpretation is potentially competing against her own. Thus, Kelly seems to be caught between satisfying the tasks of both asserting her independent understanding of the article and aligning with Tamar. On the one hand, the cut-off (i.e., "he—"), hedges (i.e., "almost like saying," the three micro pauses), and sound stretches (i.e., "breakdown," "stra:tegy," "u:se") that accompany the delivery of Kelly's counter-interpretation (lines 11-12) constitute the "delayedness" which characterize a dispreferred action (Pomerantz, 1984, p. 72). On the other hand, the same sound stretches coupled with the stresses also create a sense of deliberateness and emphasis that foreground Kelly's own distinct interpretation in the context of Tamar's. Also evidencing Kelly's dilemma is her body movement in line 13, where she noticeably looks toward the professor (who responds with a nod of approval) as she finishes her own interpretation. That is, she makes an attempt at balancing the competing needs of asserting her independent understanding and aligning with Tamar by looking to the professor, the official mediator and "knower" of the discussion, for validation of her understanding.

A more overt measure taken by Kelly to resolve the dilemma, however, lies in the second part of her response (lines 13-15) where peer referencing plays a major role. She begins with the stressed conjunct "Or," indicating that the ensuing explanation is equally feasible. Her subsequent incorporation of Tamar's conjecture is, nevertheless, parenthetically qualified with "as you were saying" (line 13). Note that Kelly accentuates the word "you," and therefore makes a point of foregrounding Tamar as the "animator," "author," and "principal" (Goffman, 1981) of the other interpretation—one that Kelly could share but would hesitate to focus

upon despite its contribution to an enriched understanding. This second part of Kelly's response is also distinctly hearable as a rush-through. The increased speed and lowered pitch (lines 13-15) contrast sharply with the earlier pauses, stresses, and sound stretches (lines 11-12). Again, Kelly makes it clear by virtue of her turn design, that although Tamar's conjecture is not far-fetched, it is not what Kelly believes to be the main thrust of the author's claim. As shown in the last line of the transcript, Kelly's stance is much more in alignment with Libby's subsequent argument that the author would encourage a holistic consideration of combined measures rather than prescribing a simplistic solution. Including Tamar's perspective while displaying Kelly's own is somewhat similar to Lerner's (1994) finding on list construction where "incorporating a prior utterance into a list of related items" is used to "achieve a qualified acceptance" in "balancing multiple social concerns" (p. 20).

Through peer referencing, Kelly acknowledges Tamar's view while withholding instant endorsement of that view. She displays her wish to align with a fellow speaker without actually and completely doing so. She makes visible not only her willingness to contemplate an interpretation which she clearly regards as secondary, but also her orientation to maintaining her "footing" (Goffman, 1981) away from that view (see also Myers, 1999 for the emphasis on detachment performed by reported speech). In short, saying "as you were saying" allows Kelly to curb, but not completely compromise her orientation to disagreeing with Tamar. That said, it is also clear from the analysis that peer referencing is only one contributing device among a host of co-occurring interactional features (e.g., prosody) which function to achieve the balance between disagreeing with a fellow member and displaying one's own comprehension.

In summary, the interactional significance of peer referencing is amenable to subtle changes in composition and locally interpretable. When the stress is placed on "you" in "as you said," peer referencing can serve to preserve the integrity of one's position without overtly disagreeing with a potentially competing one. Otherwise, it can function to either acknowledge or push the collaborative construction of a critique. Thus, compared to other argumentatively formulated "saying" expression (e.g., "So you're saying X" in Gonzales, 1996; "You say X, but what about Y?" in Hutchby, 1996), peer referencing works in much more affiliative than disaffiliative ways.

ASSERTING VULNERABILITY

I use the term *asserting vulnerability* to refer to the sorts of utterances found in the data (16 instances) where the speakers frame themselves as being vulnerably confused, uncertain, lost, not knowing, or admit that their arguments have been less than accurate, consistent, coherent, or plausible. They are composed of the first person pronoun "I" followed by expressions of vulnerability, such as "I'm really lost," "I don't know," "I wasn't sure," or "I'm really off the deep end." In an

institution that values knowledge, the paradoxical practice of essentially saying "I really don't know what I'm talking about" seems particularly interesting.

On the surface, what I call "asserting vulnerability" might bear some resemblance to hedges or disclaimers (e.g., Hewitt & Stokes, 1975; Holmes, 1982, 1984, 1995; Hyland, 1994, 1996; Skelton, 1988) in its non-assertiveness. However, asserting vulnerability differs from hedges or disclaimers in both composition and interactional significance. Hedges are linguistic devices used to express tentativeness or to "weaken or reduce the force of an utterance" (Holmes, 1995, p. 72). They mostly contain a conventional set of lexical items such as "perhaps," "would," "suggest" or pragmatic particles such as "I think" or "you know." Although Hyland (1996) calls attention to some discourse-based hedges in scientific articles such as "admission(s) to a lack of knowledge" (e.g., "We do not know whether the increase in intensity of illumination from 250 to..., p. 272"), the nature of these devices/practices is very different from vulnerability asserting utterances such as "I'm really lost" or "I don't know." The hedge "We do not know whether ...," for example, displays a cautious concern for the accuracy of what is being claimed. It is very precise about what is not known. It is content-based. In fact, in academic discourse, hedges are considered to evidence "judicious restraint and meticulous accuracy," "higher cognitive functioning," or "cognitive sophistication" (Holmes, 1995, p. 111). By contrast, vulnerability asserting utterances such as "I'm really lost" tend to sound general, less founded, and almost dramatic. More importantly, they occur in specific sequential environments where ratification is observably sought but persistently absent. In what follows, I will show that asserting vulnerability may be used to untie certain interactional deadlocks by backing down from 1) an unratified disagreement or challenge (11 instances) or 2) a successively reasserted critique (5 instances).⁵

Backdown from Unratified Disagreement

In the following segment, the group is attempting to choose one main idea based on the differing main ideas the students came up with after reading the same article. At the end of their previous meeting, the professor handed out slips of paper asking each student to write down the main idea of the article. The segment begins with the professor stating/explaining her rationale for designing the task:

[05] [1021.8]

1 Prof: [lines omitted] I want O:NE point I want people to commit (.) to one thing

2 One (.) statement of [some sort.=

3 Libby: [.hhh =When when we talk about

4 committing to one statement, do we also include (.) preknowledge of

5 genre because if you know that (.) in academic articles like Leki, she's

6 always going to make (.) a research point and a pedagogical implication point.

7 It's kind of part of (.) the genre. Therefore it's not so much one point.=It's

8 it's two points. Two interrelated points. I mean there may be more. I'm not

9 saying it just stops at two but then I

- 10 —> say- does- shall I repeat it again or does everybody get me. I'm real lost D'y
 11 get me? ((*turning to the professor*))
 12 Prof: I mean- the book that she wrote was research based but it's [continues]

Partially challenging the professor's notion of "committing to one point" in summarizing an article, in line 3 Libby goes into a multi-unit turn proposing that there could be more than one main point. Note that throughout Libby's turn at least five possible completion points are available for next speaker uptake. In fact, the turn constructional units (TCUs) following her claim that "it's two points" (line 7) are merely restatements of what has already been made clear: "Two interrelated points. I mean there may be more. I'm not saying it just stops at two but then I say-..." (lines 7-8). In these four additional TCUs, Libby continues to recycle the same point to create new transition-relevance places for next speaker uptake. When that fails, she uses a question that commands a direct answer (e.g., "Shall I repeat again or does everybody get me."). She then claims to be "really lost," and finally, directs the question "D'you get me" straight to the professor. Such is the context in which the assertion of vulnerability "I'm really lost" (line 9) is embedded.

Given the lack of any apparent inconsistency or confusion in what Libby is saying, her claim of being "lost" seems hard to justify. She might be "lost" in terms of not hearing any feedback, receiving any uptake, or gaining any agreement from the group throughout her multi-unit turn, but not so with reference to developing her argument. Put otherwise, Libby's assertion of vulnerability has nothing to do with the substance of her claim, that is, the co-existence of multiple main points in one article. In fact, her claim is rather forcefully advanced. For example, she not only says that an author like Leki often makes two points (line 5-6), but also legitimizes the practice by pointing out that making two points is a requirement of the genre (line 7). Later on, she further upgrades her claim by saying "I mean there may be more. I'm not saying it just stops at two" (lines 8-9). What is observably lacking, though, is any uptake following her challenge of the professor either from the professor herself or the rest of the members. By asserting vulnerability at this moment of interactional deadlock, Libby shifts her role from one who is aggressively advancing a challenge to one in need of guidance and reassurance. She redesigns her turn so that it calls for advice rather than a second assessment (i.e., agreement or disagreement). Although what eventually brings about the professor's uptake are undoubtedly the two questions (in line 10-11), or more precisely, the last question "D'you get me?" (cf. "current speaker selects next" in Sacks et al., 1974), what fundamentally changes the dynamics of the interaction, or rescues the interaction from the narrow strait it has come to, appears to be Libby's backdown remark "I'm really lost."

Although to my knowledge the term "backdown" has not been explicitly defined in previous literature, its use tends to involve some sort of reversal or revision of one's earlier position. In Coulter's (1990) study on argument sequences, for example, he characterizes a backdown as a third position action after an ex-

change of assertion and counter-assertion, and calls attention to “backdown-implicative silences” (e.g., silence after a counter-assertion) and “explicit backdowns” (e.g., the utterance “I know” after a sequence of accusation and denial). Pomerantz (1984) also discusses a type of backing down in the case of potential disagreement:

- [06] SBL: 3.1-8 (from Pomerantz, 1984, p. 76)
 B: ...an' that's not an awful lotta fruitcake.
 (1.0)
 —> B: Couse it is. A little piece goes a long way.
- [07] JS: II: 48 (from Pomerantz, 1984, p. 76)
 L: D'they have a good cook there?
 (1.7)
 —> L: Nothing special?

Compared to these examples, “I’m really lost” does not contain any explicit reversal or revision of position. Thus, I use the term backdown somewhat differently from the way it has been used in previous literature. Although Libby renders questionable an important basis upon which her previous assertions were made by characterizing her mindset as being “really lost,” without any explicit indication of position reversal or revision, her backdown is more symbolic than substantive. What is being reversed is her role from a “challenger” to a “vulnerable student” in search of guidance and reassurance. Asserting vulnerability is therefore a backdown tactic exercised to exit an interactional deadlock created by unratified disagreement. It is Libby’s solution for resolving the competing agendas between asserting her independent understanding on the one hand, and seeking approval from the professor and the rest of the group on the other.

Backdown from Successively Reasserted Critique

Asserting vulnerability is also used as a strategic backdown when one’s critique of an article persistently fails to induce any endorsement from the group. For example, leading up to the following segment has been Jack’s argument that strategy training is necessary, a point at odds with what seems to be suggested in the assigned article. Meanwhile, this critique of Jack’s has been repeatedly countered by the professor’s attempt to render Leki’s point acceptable. The professor first called attention to Leki’s argument that a simplified text designed to teach how to read (i.e., do strategy training) would not help one develop the ability to eventually read authentic materials. She then interpreted Leki’s point as suggesting that strategy training is not enough because a reader’s ultimate goal is not to read, but to get information out of the reading materials. Jack, however, insisted that the ultimate goal is a distant one while learning reading strategies is a necessary, intermediate step. Immediately before the segment below, the professor pointed out that the

gist of Leki's critique of strategy training has to do with her concern that books often stop at teaching isolated strategies without addressing when, where, and how the strategies can be used to get information from reading:

[06] [1021.20]

1 Prof: [omitted talk] Ultimately don't you want to be able to use these?

2 (0.2)

3 Jack: Right. [M. I-

4 Prof: [I think.

5 Jack: I agree with her um strategy choice selection is that something that good readers
6 —> are able to deal with. Not something that () not as good, but (.) I don't know. I
7 just thought- I wasn't sure either whether she was saying that strategy,
8 teaching strategy (is) um a good thing? I thought she was saying that it
9 was (.) unnecessary.

10 Prof: Is she saying that?

Jack's seemingly veiled critique is first of all enfolded in the "I agree with her...but" structure (lines 5-6). It is also repeatedly alluded to in the talk prior to the segment. At one point earlier in the discussion, Jack said, "there are strategies that I have to develop (.) um to read magazines and newspapers in this language." Ten turns later, he expressed the same concern again in a much more direct tone: "they [Leki's points] just seemed sorta- it's so anti-pedagogical to me." He persisted eight turns after that: "but I think it [strategy training] does help you because even in simplified texts, there are strategies that are going to be um (.) highlighted (.) that can identify to the (.) authentic text." And he pressed on yet again later: "but don't you need to be able to scan? to to um attain this (.) higher, you know, more interesting goal?"

In the above segment, Jack insinuates, for the last time, his dissatisfaction with what seems to him to be the deemphasis on strategy training in Leki's article. At this point, however, his dissatisfaction is packaged in much more tentative talk. In lines 5-6 he begins by agreeing with Leki's contention that being able to use strategies appropriately is a characteristic of good readers—a point alluded to in the professor's prior turn. However, Jack's turn construction also shows that he does not treat the issue of appropriate strategy choice as intimately relevant to his central concern—the value of strategy training. The agreement which ends in non-final intonation is immediately followed by the contrastive marker "but" and a micro pause. But instead of spelling out his critiquing position as he has repeatedly done earlier, Jack ends the main clause of his TCU with a rather general assertion of vulnerability: "I don't know" (Line 6). He then proceeds to elaborate upon the "I don't know" with "I just thought," which he cuts-off with a self-initiated repair (Schegloff, Jefferson & Sacks, 1977) that transforms it into another assertion of vulnerability: "I wasn't sure" (Lines 6-7). That is, Jack replaces the originally planned "I just thought she was saying that it [strategy training] was unnecessary" with "I wasn't sure either whether she was saying that..." In other

words, despite Jack's apparent refusal to accept the importance of appropriate strategy selection as the reason for de-emphasizing strategy training, he employs interactional practices to tone down his original assertion.

Again, Jack is certainly not unsure about his position vis-a-vis the reading. He wouldn't have repeatedly voiced the same thesis had that been the case. This, in fact, resonates with my earlier observation regarding Libby's assertion of being "really lost" when she was clearly adamant about where she was heading. The nature of vulnerability in both cases seems more germane to the speakers' uncertainty about whether their observations are being endorsed. In both cases, these assertions of vulnerability occur in sequences where ratification is observably sought but persistently absent. They are both positioned alongside clearly formulated arguments, and by virtue of this juxtaposition, they render their arguments less presumptuous, confrontational, or uncompromising, and thereby signal an interactional backdown.

This particular use of asserting vulnerability is also observable in the following segment. Seven minutes before the segment, Libby started arguing against the notion that L2 readers have a unidimensional approach toward lexicon while L1 readers are capable of deriving various meanings of the same word from the context. Despite the various attempts made by the professor and other members to interpret this claim from different angles for the past seven minutes, Libby insists on questioning its validity by arguing that the readers' linguistic proficiency or literacy level needs to be taken into account. At the arrow below, however, Libby stops from yet another attempt to persist, and instead, switches to an assertion of vulnerability:

[07] [1007.41]

- 1 Kelly: [When I was reading a:nd when I- oh, when I was reading, I >looked at it
 - 2 whatever context it was in and I [thought of it ()]
 - 3 Libby: [But that's my point.] and if the context that
 - 4 —> you have- ∞you know what? I think- I'm [making us all=
 - 5 Tamar: [hhh
 - 6 Libby: beat[ing a dead horse°.
 - 7 Tamar: [>°yeah°< no no no you're right. but the context is
- in the text, whereas the uh the list of all the meanings of bank is in your mi:nd.

By saying "I'm making us all beating a dead horse," Libby evaluates the activities she has been engaged in, or more specifically, her insistent attempts of critiques, as useless or without consequence.⁶ In so doing, she signals a backdown from her repeatedly advanced critique. The conceding tone of backing down is first instantiated in the markedly lowered pitch in which the assertion of vulnerability is delivered (lines 4 and 6). Backing down as an action accomplished in asserting vulnerability is also evidenced in the next turn where Tamar enthusiastically says "°yeah° no no no you're right" in overlap (line 7), supporting Coulter's (1990) claim that "further counter-assertions are dispreferred" after a backdown has been

produced. In addition, Tamar's affiliative uptake also confirms Pomerantz' (1984) analysis that disagreement is preferred after self-deprecation. Again, asserting vulnerability is used to terminate successive failures to achieve negotiated consensus. Meanwhile, the assertion projects a sharp contrast to the persisting position Libby has been taking, thereby mitigating its force, and re-orienting the dynamics around the table.

In sum, the device of asserting vulnerability is embedded within a sequence where the speaker's forcefully articulated disagreement or persistently advanced critique fails to occasion clear concurrence from the others. The reference to one's own inadequacy entailed in the assertion is often more interactionally motivated than substantive. It might be argued that this practice is indirectly linked to the "self-praise avoidance" constraint identified in Pomerantz (1978, p. 88). It serves to untie the communicative deadlock by recasting oneself as being less aggressive and uncompromising, expanding the possibilities of uptake, and terminating a sequence of talk in which consensus is clearly unlikely. In that sense, it is a resolution that works to strike a balance between asserting one's independent understanding and seeking alignment with the group.

In the conversation/discourse analytic literature, other practices found to "resolve" an interactional impasse include self-selection by a third party (Gonzales, 1996, p. 92) and leaving the scene (M. H. Goodwin, 1990, p. 215). According to Kotthoff (1993), concessions are dispreferred once an initial disagreement sequence is produced, and they are delivered with reluctance markers in a stepwise fashion. In fact, "unprepared position shifts can be regarded by the interlocutors as the inability to defend an opinion" (Kotthoff, 1993, p. 193). Given this background, abruptly backing down from clearly stated assertions appears to characterize graduate students' transitional stage between undergraduate novicehood and Ph.D level junior expertise.⁷ In a sense, it symbolizes the pendulum swinging from being a junior expert proclaiming independent viewpoints to an undergraduate novice begging guidance and reassurance. Finally, it is worthy of note that the consensus towards which asserting vulnerability is oriented is in broad consonance with the preference for agreement (Sacks, 1987), the preference for using explicit backdowns to "terminate an argument sequence" (Coulter, 1990, p. 181) found in ordinary conversation, and more generally, the preference for maintaining social solidarity (Heritage, 1984; Pomerantz, 1984).⁸

CONCLUSION

Broadly speaking, the types of practices dealt with in this paper concern what scholars in communication and education studies recognize as the social affective domain of group interaction, which is intricately related to what can be achieved in a group (e.g., Barnes & Todd, 1995; Friend & Cook, 1992). The intricate relationship between a group's social affective practice and its substan-

tive achievements finds its unique expression in a graduate seminar, especially in the doing of disagreement and critique.

In this study, I have tried to detail the composition, position, and action of two resources employed by the participants to cope with the tricky task of disagreeing and critiquing. Depending on the subtle changes in composition and the type of interactional task at hand, peer referencing (e.g., "as you said") is used to include another's view while withholding an affiliative response and preserving the integrity and continuity of one's intellectual stance. It can also work as a building-block or an evidential piece in supporting one's challenge of the professor or critique of published articles. Asserting vulnerability (e.g., "I'm really lost") is deployed in sequential environments where one member's persistently and forcefully displayed disagreement, challenge or critique has brought the discussion to an interactional deadlock. In asserting vulnerability, the speaker makes an attempt to break the impasse by redefining him/herself as being receptive to advice rather than aggressive and resistant to negotiation.

I have demonstrated how peer referencing and asserting vulnerability operate to manoeuvre the tensions between learning and assessing, and between being a receptive, abiding learner and an assertive, independent thinker. To a certain extent, the various actions accomplished in peer referencing and asserting vulnerability combine to delineate the sense of "in-betweenness" of being a graduate student in a reading seminar. Their assertion of independently formulated positions often seems complicated by the competing interest in seeking approval of these positions. This sense of "in-betweenness" is in part captured in using peer referencing to acknowledge the validity of another's view while advancing one's own, potentially conflicting position. It is also observable when peer referencing is used to twist an otherwise independently formulated critique into a co-constructed one. Finally, it is manifested in the abrupt backdown, via the assertion of vulnerability, from clearly and persistently articulated disagreement or critique.

The discussion in this paper is by no means an exhaustive account of the discourse resources available to the participants in balancing the competing agendas in seminar discussions. In particular, the analysis in this study is based upon a small number of instances in a single seminar. Future research needs to evaluate the validity of these practices with a larger collection of instances in various seminars across the curriculum. A collection of instances where the same practice is used across different seminars will allow access to a deeper understanding of how a particular strategy works. Overall, it is hoped that this study has provided a glimpse into the ways in which the participants of a graduate seminar managed the complex tasks of disagreeing and critiquing. It is also hoped that the findings of this study will serve as a potentially valuable resource for the growing efforts to include discussion skills in the design of pedagogical activities.

ACKNOWLEDGEMENTS

Some of the materials presented here were submitted as part of my doctoral dissertation completed under the thoughtful guidance of Leslie Beebe and Jim Purpura at Teacher's College, Columbia University in May 2000. I thank Anne Ediger and her eight seminar members for their generosity and support in making the data available. I gratefully acknowledge the probing and constructive comment provided by the two anonymous readers. Leah Wingard, the editor of *Issues in Applied Linguistics*, has also offered facilitating directions in the revision of this paper. Finally, Michael Waring has been a patient and helpful sounding board throughout the many drafts.

APPENDIX TRANSCRIPTION SYMBOLS

(.)	untimed perceptible pause within a turn
<u>underline</u>	stress
CAPS	very emphatic stress
↑	high pitch on word
.	sentence-final falling intonation
?	yes/no question rising intonation
,	phrase-final intonation (more to come)
-	a glottal stop, or abrupt cutting off of sound
:	lengthened vowel sound (extra colons indicate greater lengthening)
=	latch
→	highlights point of analysis
[]	overlapped talk
[]	
°soft°	spoken softly/decreased volume
> <	increased speed
()	transcription impossible
(words)	uncertain transcription
[]	comments on background or skipped talk
(())	non-speech activity such as laughter

(adapted from Atkinson and Heritage, 1984)

NOTES

¹ I owe this clarification of focus to an anonymous reviewer and Leah Wingard.

² The wording used here and on several occasions after this (i.e. "undergraduate novicehood" and doctoral level "junior expertise") was suggested by an anonymous reviewer.

³ These are pseudonyms.

⁴ Sacks notes such preferences in questions like "Ken you walk?" or "Ud be too hard you yuh?" (Sacks, 1987, p. 64)

⁵ It needs to be pointed out that I use the term "backdown" in a sense that is more symbolic than substantive (see discussion next section).

⁶ I thank Leah Wingard for suggesting this elaboration.

⁷ I owe this insight to an anonymous reviewer.

⁸ Here I have taken the liberty of using terminology that is more "interactant's world"-based (Pomerantz, 1989). In fact, Pomerantz (1989) calls for "moving back and forth between different analytic approaches" in enriching our understanding of an interactional phenomenon (p. 246). In her

attempt to translate a sequence-focused analysis to an "interactant's" world analysis," Pomerantz (1989) refers to "preference for agreement" as "proper," "satisfying," or "comfortable" (p. 245), which contrasts with the sequentially based notion of preference where preferred actions are performed straightforwardly without delay while dispreferred actions are delayed and qualified (Heritage, 1984; Levinson, 1983). Similarly, Heritage (1984) also points out that "preferred actions are generally supportive of social solidarity" (p. 269), and that "issues of 'face'" are closely related to the organization of talk (p. 268). According to Pomerantz (1989), the exercise of translation is a fruitful endeavor that allows us to develop a fuller understanding of the different aspects of the same phenomenon (see also Viechnicki, 1997 on bridging the gap between Conversation Analysis and Goffman's work).

REFERENCES

- Atkinson, J. M., & Heritage, J. (1984). *Structures of social action: Studies in conversation analysis*. Cambridge: Cambridge University Press.
- Barnes, D., & Todd, T. (1995). *Communication and learning revisited: Making meaning through talk*. Portsmouth, NH: Boynton/Cook Publishers.
- Brzozko-Barratt, K., & Johnson-Saylor, K. (2001,). *Talk and interaction in the graduate seminar*. Paper presented at American Association for Applied Linguistics Annual Conference, St. Louis, MO.
- Coulter, J. (1990). Elementary properties of argument sequences. In G. Psathas (Ed.), *Interaction competence* (pp. 181-203). Washington, D. C.: University Press of America.
- Davidson, J. (1984). Subsequent versions of invitations, offers, requests, and proposals dealing with potential or actual rejection. In J. M. Atkinson & J. Heritage (Eds.), *Structures of social action* (pp. 102-128). Cambridge: Cambridge University Press.
- Ferris, D. (1998). Students' views of academic aural/oral skills: A comparative needs analysis. *TESOL Quarterly*, 32, 289-315.
- Friend, M., & Cook, L. (1992). *Interactions: Collaboration skills for school professionals*. New York and London: Longman.
- Goffman, E. (1981). *Forms of talk*. Oxford: Blackwell.
- Gonzales, P. A. (1996). *The talk and social organization of problem-solving activities among physicists*. Unpublished doctoral dissertation, University of California, Los Angeles.
- Goodwin, C. (1981). *Conversational organization: Interaction between speakers and hearers*. New York: Academic Press.
- Goodwin, C. (1984). Notes on story structure and the organization of participation. In J. M. Atkinson and J. Heritage (Eds.), *Structures of social action* (pp. 225-246). Cambridge: Cambridge University Press.
- Goodwin, M. H. (1990). *He-said-she-said: Talk as social organization among black children*. Bloomington and Indianapolis: Indiana University Press.
- Grimshaw, A. D. (1989). *Collegial discourse: Professional conversation among peers*. Norwood, NJ: Ablex Publishing Corporation.
- Hartmann, P., & Blass, L. (2000). *Quest: Listening and speaking in the academic world*. New York: McGraw Hill.
- Hemmert, A., & O'Connell, G. (1998). *Communicating on campus*. San Francisco: Alta Book Center Publishers.
- Heritage, J. (1984). *Garfinkel and ethnomethodology*. Cambridge: Polity Press.

- Heritage, J. (1985). Analyzing news interviews: Aspects of the production of talk for an overhearing audience. In T. van Dijk (Ed.), *Handbook of discourse analysis* (Vol. 3, pp. 95-119). London: Academic Press.
- Heritage, J. C., & Watson, D. R. (1979). Formulations as conversational objects. In G. Psathas (Ed.), *Everyday language: Studies in ethnomethodology* (pp. 123-162). New York: Irvington Publishers.
- Hewitt, J. P., & Stokes, R. (1975). Disclaimers. *American Sociological Review*, 40 (1), 1-11.
- Holmes, J. (1982). Expressing doubt and certainty in English. *RELIC Journal*, 13 (2), 9-28.
- Holmes, J. (1984). Modifying illocutionary force. *Journal of Pragmatics*, 8, 345-365.
- Holmes, J. (1995). *Women, men and politeness*. London: Longman.
- Holt, E. (1996). Reporting on talk: The use of direct reported speech in conversation. *Research on Language and Social Interaction*, 29, 219-245.
- Holt, E. (2000). Reporting and reacting: Concurrent responses to reported speech. *Research on Language and Social Interaction*, 33, 425-454.
- Hutchby, I. (1996). *Confrontation talk: Arguments, asymmetries, and power on talk radio*. Mahwah, NJ: Erlbaum.
- Hyland, K. (1994). Hedging in academic writing and EAP textbooks. *English for Specific Purposes*, 13(3), 239-256.
- Hyland, K. (1996). Talking to the academy. *Written Communication*, 13(2), 251-281.
- Jacoby, S. (1998). *Science as performance: Socializing scientific discourse through the conference talk rehearsal*. Unpublished doctoral dissertation, University of California, Los Angeles.
- Jacoby, S., & Gonzales, P. (1991). The constitution of expert-novice in scientific discourse. *Issues in Applied Linguistics*, 2, 149-181.
- Jacoby, S., & McNamara, T. (1999). Locating competence. *English for Specific Purposes*, 18(3), 213-241.
- Johns, C. M., & Johns, T. F. (1977). Seminar discussion strategies. In A. P. Cowie & J. B. Heaton (Eds.), *English for academic purposes* (pp. 99-107). BAAL/SELMOUS Publications.
- Jones, J. F. (1999). From silence to talk: Cross-cultural ideas on students' participation in academic group discussion. *English for Specific Purposes*, 18(3), 243-159.
- Kotthoff, H. (1993). Disagreement and concession in disputes: On the context sensitivity of preference structures. *Language in Society*, 22, 193-216.
- Lakoff, R. (1990). *Talking power: The politics of language*. New York: Basic Books.
- Leki, I. (1992). *Understanding ESL writers: A guide for teachers*. Portsmouth, NH: Boynton/Cook Publishers.
- Lerner, G. (1994). Responsive list construction: A conversational resource for accomplishing multifaceted social action. *Journal of Language and Social Psychology*, 13(1), 20-33.
- Levinson, S. (1983). *Pragmatics*. Cambridge: Cambridge University Press.
- Li, C. N. (1986). Direct and indirect speech: A functional study. In F. Coulmas (Ed.), *Direct and indirect speech* (pp. 29-45). Berlin: Mouton.
- Lynch, T., & Anderson, K. (1992). *Study speaking*. Cambridge: Cambridge University Press.
- Myers, G. (1999). Functions of reported speech in group discussions. *Applied Linguistics*, 20, 376-401.

- Ochs, E., & Jacoby, S. (1997). Down to the wire: The cultural clock of physicists and the discourse of consensus. *Language in Society*, 26, 479-505.
- Pomerantz, A. (1978). Compliment responses: Notes on the co-operations of multiple constraints. In J. Schenkein (Ed.), *Studies in the organization of conversational interaction* (pp. 79-112). New York: Academic Press.
- Pomerantz, A. (1984). Agreeing and disagreeing with assessments: Some features of preferred/dispreferred turn shapes. In M. Atkinson and J. Heritage (Eds.), *Structures of social action* (pp. 57-101). Cambridge: Cambridge University Press.
- Pomerantz, A. (1989). Epilogue. *Western Journal of Speech Communication*, 53, 242-246.
- Price, J. (1977). Study skills—with special reference to seminar strategies and one aspect of academic writing. In S. Holden (Ed.), *English for specific purposes* (pp. 26-29). Modern English Publications Limited.
- Prior, P. (1998). *Writing/disciplinarity: A sociohistoric account of literate activity in the academy*. Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.
- Psathas, G. (1995). *Conversation analysis: The study of talk-in-interaction*. Thousand Oaks, CA: Sage Publications.
- Raimes, A. (1999). *Exploring through writing: Academic writing in a second language*. Cambridge: Cambridge University Press.
- Sacks, H. (1987). On the preferences for agreement and contiguity in sequences in conversation. In G. Button & J. R. E. Lee (Eds.), *Talk and social organization* (pp. 54-69). Clevedon: Multilingual Matters.
- Sacks, H., Schegloff, E. A., & Jefferson, G. (1974). A simplest systematics for the organization of turn-taking in conversation. *Language*, 50, 696-735.
- Schegloff, E., Jefferson, G., & Sacks, H. (1977). The preference for self-correction in the organization of repair in conversation. *Language*, 53, 361-82.
- Skelton, J. (1988). The care and maintenance of hedges. *ELT Journal*, 42 (1), 37-43.
- Steer, J. (1995). *Strategies for academic communication*. Boston, MA: Heinle and Heinle.
- Tracy, K. (1997). *Colloquium: Dilemmas of academic discourse*. Norwood, NJ: Ablex Publishing Corporation.
- Tracy, K., & Baratz, S. (1993). Intellectual discussion in the academy as situated discourse. *Communication Monographs*, 60 (4), 301-320.
- Tracy, K., & Carjuzaa, J. (1993). Identity enactment in intellectual discussion. *Journal of Language and Social Psychology*, 2, 171-194.
- Viechnicki, G. B. (1997). An empirical analysis of participant intentions: Discourse in a graduate seminar. *Language and Communication*, 17(2), 103-131.
- Waring, H. Z. (in press). Displaying substantive reciprocity in seminar discussion. *Research on language and social interaction*.

Hansun Zhang Waring received her doctorate in Applied Linguistics from Teacher's College, Columbia University in May 2000. She is currently an Assistant Professor of Speech Communication at Mercy College in New York. Her research interests are discourse analysis (conversation analysis), interlanguage pragmatics, and second language acquisition.

Cohesion and Coherence in Children's Written English: Immersion and English-only Classes

Jungok Bae

University of California, Los Angeles

This study investigates the nature of cohesion, coherence, content, and grammar emergent in children's essays, with a greater emphasis given to the understanding of cohesion and coherence. Conceptual definitions of these constructs are summarized based on prior research. The measurement of these constructs is operationalized into a picture-based narrative writing task for elicitation and scoring criteria for quantification. 192 first and second graders from an immersion program and English-only classes participated in the study. The analysis uses percentages, correlations, multiple regression, and qualitative analyses. Main findings include the following: (a) the measurement of cohesion and coherence can be operationalized; (b) referential and lexical cohesion correlate highly with the overall writing quality defined as the sum of the ratings of coherence, content, and grammar; (c) ellipses and substitution show a weak correlation with the overall writing quality; (d) lexical and referential cohesion are significant predictors of coherence while other types of cohesion are not; (e) dominant reference types are pronominal forms and proper nouns, and prominent types of conjunctive relation are temporal and additive; and (f) the most common error in cohesion is inaccurate reference. The substance and method of this study can provide a foundation for investigating subsequent topics with latent variables and different linguistic backgrounds and grade levels.

This study aims to understand certain linguistic and semantic resources for text construction, namely the constructs of cohesion, coherence, content, grammar, and text length in English writing. A greater emphasis in this study is given to the understanding of cohesion and coherence. With this in mind, their relations to other salient writing constructs, such as content, grammar, and text length are investigated.

Generally, the concepts of cohesion and coherence are more technical and relatively uncommon to many people compared to the concepts of other more universally understood language-related components, such as grammar, content, and text length. One of the most significant works that have contributed to our explicit understanding of cohesion is Halliday and Hasan (1976/1993). According to Halliday and Hasan (1976/1993, p. 4), the concept of cohesion is a semantic one, referring to "relations of meaning" that exist within the text, and it "occurs where the interpretation of some element in the discourse is dependent on that of another." Cohesion is expressed "partly through the grammar and partly through the vocabulary" (p. 5). In comparison, coherence, generally defined, refers to the quality of a text when it makes sense or is pleasing because all the parts or steps fit

together well and logically (Collins Cobuild, 1996). It is the connection that is established partly through cohesion (Halliday & Hasan, 1989) and partly through something outside the text that is usually the knowledge which a listener or reader is assumed to possess (Renkema, 1993, p. 35), such as background knowledge, genre expectations, and reader expectations. Extensive studies are available that offer theoretical discussions of cohesion and coherence (Bamberg, 1984; Brown & Yule, 1983; Connor & Johns, 1990; Cook, 1989; Gundel, Hedberg, & Zacharski, 1993; Halliday, 1985; Halliday & Hasan, 1976/1993, 1989; Koshik, 1999; McCarthy, 1991; Oller & Jonz, 1994; Renkema, 1993; Smith, 1984).

Grounded in current theories of cohesion and coherence, this paper outlines theoretical definitions of cohesion, coherence, content, and grammar. The study then operationalizes the measurement of the postulated constructs of coherence and cohesion, as well as content, grammar, and text length as a first step toward developing assessment measures of the constructs. Specifically, the theoretical constructs are realized into operational definitions of the constructs by means of a narrative writing task and scoring criteria used to elicit and quantify the constructs. This study is essentially about *construct validation* (or to use Nunnally's 1978 alternative term for construct validation suitable to this paper, *construct explication*), which is the process of elaborating and refining the meaning of the constructs on the basis of empirical evidence. In sum, cohesion and coherence, which have been given ample theoretical discussions in prior research, are given an empirical investigation and evidence based on assessment procedures in this study. This measurement-based empirical augmentation is what is different from prior research in the theory of cohesion and coherence.

Context of the Study

This investigation of cohesion, coherence, content, and grammar was undertaken within the context of a two-way immersion program, the Korean/English Two-Way Immersion Program (Campbell et al., 1994; Kim, 1996; Walker, 1992), and English-medium classes. The immersion program consists largely of Korean-language background and English-language background students, and the English-medium classes consist of students who are proficient in terms of English oral skills. Although the present study is not intended to compare writing in these instructional programs (see below), the student populations are of interest in that immersion classes and English-medium classes have rarely been used in the study of cohesion and coherence. The English level of all the students in the English classes and the majority of students from the immersion program was identified as proficient in terms of oral English skills at the time of the study. However, the levels of literacy skills of these children differ widely even among the students who have a fluent command of oral English. Using the additional data from Korean-background students who were initially identified as Limited English Proficient (LEP) but who were becoming proficient bilinguals in the immersion pro-

gram at the time of the study, interesting aspects of the specified constructs on the learning curve could be incorporated in the study.

Purpose of Paper

In the present paper, immersion classes and English-medium classes are grouped together in studying the nature of the specified constructs. The concern in this paper is not which instructional program performs better or whether distinct or same characteristics of text qualities are found between these programs. Rather, by incorporating the two instructional groups, the paper aims to substantiate robust constructs of English writing that reflect the text characteristics across these programs. It can be said that, in the present paper, the robust English writing constructs are established as a basis for evaluating the learning of English writing that is going on and for comparing learning across programs in subsequent studies.

Research Questions

This study addresses the following research questions:

1. What characteristics of cohesion and coherence are observed in the narrative writing of children (including immersion students and students in English-only classes), and how are these two qualities related to each other and to other salient components of language such as content, grammar, and text length?
2. Which types of cohesion are more or less prominent in these children's narrative writing?
3. What are the patterns of errors in cohesion for English-language background children and Korean-language background children?

Significance and Rationale

This section discusses the significance of the approaches used in this study. The significance is addressed in the areas of internal and external aspects of construct explication, rationale for selecting the constructs, and the operationalization of the constructs.

Internal and external aspects of investigation

Extensive prior research that focuses on cohesion specifically is available (Cook, 1989; Cox, Shanahan, & Sulzby, 1990; Freedle, 1991; Halliday & Hasan, 1976/1993; Lindsay, 1984; Norris & Bruning, 1988; Ricento, 1987). Very few studies, if any, have looked into cohesion in relation to constructs external to cohesion especially at the statistical level. While this study gives a greater emphasis to the internal aspect of cohesion, it also introduces external aspects of investigations into cohesion. Specifically, this study examines not only the definition of cohesion and its subdimensions, but also illuminates cohesion in relation to other language-related variables external to cohesion, such as coherence, content, grammar, and text length. Ideally, an unlimited number of linguistic and nonlinguistic variables could be included as external variables in a validation of target con-

structs; each external variable would contribute to our understanding of the focal constructs. In practice, however, it is impossible to include all potential variables in one study. Thus, in this study coherence, content, grammar, and text length have been selected since these variables are all salient components in written texts. Coherence receives special interest because coherence has often been compared to and distinguished from cohesion (e.g., Carrell, 1982; Cox, Shanahan, & Tinzmann, 1991; Enkvist, 1990; Fitzgerald & Spiegel, 1986; Koshik, 1999; Oller & Jonz, 1994; Spiegel & Fitzgerald, 1990). Grammar and content are traditionally considered important objectives in the teaching and testing of writing: Grammar represents a linguistic domain of language and content represents the semantic domain of language. It is of interest to examine the extent to which the two constructs of linguistic and semantic properties are related to coherence and cohesion. In addition, text length is often attended to by readers and writers, and it would be of interest to see if text length has something to do with ability in the other constructs. Since we can easily quantify text length by simply counting words, text length is included in examining its relation to the above constructs. These constructs come into play in the analysis, contributing to the understanding of one another as external variables.

Operationalization of the measurement of cohesion and coherence

The significance of this study also concerns methodology employed for operationalization. In particular, number of subjects, elicitation method, and quantifications used in the process of operationalization are noteworthy.

With regard to N (number of subjects), research that analyzes written text qualities typically uses case studies, which usually focus on descriptive, qualitative analysis of one or several study participants. With the N of a case study, statistical analysis becomes very limited. For instance, correlations, ANOVA, MANOVA, and regression analyses require a large number of subjects—although precisely how large is “large” is debatable—to apply inferential statistics involving statistical significance tests in generalizing findings. This study uses data collected from several elementary school classes (a total of 192 students). With this N , statistical analyses such as correlations, regression, and percentage calculations can be performed. This resource thus contributes to the power of the research by enhancing the reliability and generalizability of findings about the text qualities in a way that has not been possible in prior research on cohesion and coherence that use case study analyses.

Secondly, methods for eliciting and quantifying the writing constructs introduced in this study, consisting of a writing task and scoring criteria, are notable. Most studies that use children’s data have used spoken data. Consequently, there is a dearth of written data collected from children. This is regrettable because children’s writing can provide a promising area of research in language acquisition and assessment. For instance, it is easier to observe language use in writing than in speaking and listening. Children as learners of literacy skills also reveal

interesting developmental features in their writing. The lack of research using children's writing is also due to the fact that young children develop oral skills before they are exposed to written materials. Due to this trend, writing assessment for children has only been in its preliminary stage, leading to a paucity of methods designed to elicit and assess writing samples from them, which, in turn, results in the lack of research on children's written data. In addition, during recent decades, educators have been aware of the limitation of multiple-choice testing and assiduously called for performance-based assessment to reform education (Arter & Spandel, 1992; Baker, O'Neil & Linn, 1993; Herman, Aschbacher & Winters, 1992; Linn, Baker & Dunbar, 1991; Mehrens, 1992). To date, measures of productive language skills such as writing have not been available as part of standardized tests due to the demands of large-scale testing, which favors multiple-choice test items. This study employs a performance-based task based on an original picture series via group testing in which students are required to produce written texts in the form of stories. This procedure makes the task not only appropriate for children but also educationally beneficial. The task with the writing prompt can also provide a potential resource for future research in children's writing.

In regards to quantification, extensive scoring rubrics are available for measuring grammar and content. Organization, which may be considered similar to or part of coherence, is often scored based on scoring rubrics in assessment projects. I know of no studies to date that have offered a quantification procedure for coherence. This study takes a proactive stance to formulate rating criteria for measuring coherence based on our theoretical understanding of coherence. It also introduces a simple scoring method of counting markers of cohesion to assess the extent to which cohesion is expressed in children's compositions; this quantification method is designed to be sensitive to the unique characteristics of cohesion, that is, distinct subdimensions and overt markers of cohesion, as we shall see below.

Construct Definitions

The constructs (or ability components) being investigated in this study are cohesion and coherence emerging in children's narrative storywriting. These two components will receive special consideration and elaboration in providing definitions. In addition, definitions of content, grammar, and text length are also provided.

Cohesion

Cohesion refers to the range of grammatical and lexical possibilities that exist for linking an element of language with what has gone before or what follows in a text: This linking is achieved through relations in meaning that exist within and across sentences (Halliday & Hasan, 1976/1993, p. 10, 33). Cohesion is confined to the specific, micro-local level of organization between and within individual clauses, thus creating connections between parts. Beginning with Halliday and Hasan, many researchers have identified several types of cohesion. The types

included in this study are reference, lexical ties, conjunction, ellipsis, and substitution, which have been traditional topics in theories of cohesion (Brown & Yule, 1983; Cook, 1989; Halliday, 1985; Halliday & Hasan, 1976, 1989; Renkema, 1993). Definitions and examples of these types of cohesion are presented in Table 1 below.

Coherence

Table 1: Types of Cohesion*

REFERENCE: Items that refer to something else in the text for their interpretation.

- (1) pronominal: e.g., *he, her, they, theirs*
- (2) proper nouns: e.g., *Brandon, Ms. Sharon*
- (3) demonstratives: e.g., *this/these, that/those, here/there*
- (4) comparatives: identity/similarity/difference/ordinals/comparatives/superlatives, e.g., *same, similar, such, bigger*

CONJUNCTION: Connectors between two independent sentences.

- (1) Additive: e.g., *and, or, by the way*
- (2) Adversative: e.g., *but, yet, however, rather*
- (3) Causal: e.g., *so, therefore, thus*
- (4) Temporal: e.g., *and then, then, after that, soon, finally*

ELLIPSIS: Elements left unsaid or unwritten but are understood by the reader/speaker.

- (1) Noun ellipsis: delete nouns, e.g., *He liked the blue hat; I myself liked the white.*
- (2) Verbal ellipsis: delete verbs, e.g., *Tom drew a small boat and April a big boat.*
- (3) Clausal ellipsis: delete clauses, e.g., A: *Will you go?* B: Yes.
A: *Would you like something to drink?* B: Sure.

SUBSTITUTION: The replacement of word or structure by a "dummy" word.

- (1) Noun substitution: e.g., *Tom drew a big boat and April drew a small one.*
- (2) Verb substitution: e.g., *He wanted to draw pictures there, and they really did.*

LEXICAL TIES:

- (1) Collocation: e.g., *go home, have fun, rain/rainy/wet/umbrella/soaked*
- (2) Repetition: e.g., *drew/draw/drawing, rain/raining/rainy*
- (3) Synonym: e.g., *sad/unhappy*
- (4) Antonym: e.g., *boy/girl, big/small*
- (5) Hyponymy (general-specific relations): e.g., *fruit/banana, apple*
- (6) Meronymy (part-whole relations): e.g., *house/door, room, wall, kitchen*

* Summarized from Cook, 1989; Halliday & Hasan, 1976, 1989; McCarthy, 1991; Renkema, 1993.

Coherence is a plot-motivated overall structure (in narrative) or plan on the macro level (Berman & Slobin, 1994, p. 67). It is an overall discourse-level property that makes a text hold together (Fitzgerald & Spiegel, 1990, p. 263).

Coherence, according Halliday and Hasan (1989), can be created by cohesive markers that are appropriately used. Halliday and Hasan (p. 95) comment that early discourse of students in a new field is relatively less coherent than their later discourse because the semantic relations between the key concepts, that is,

cohesion, are not yet clear (see also Halliday & Hasan, 1976/1993, p. 4, 8). Cohesive markers alone, however, do not necessarily make the text coherent and comprehensible. A text full of cohesive markers that are locally correct could be incoherent and incomprehensible as a whole (Oller & Jonz, 1994) as in the following example from Enkvist (1990):

My car is black. Black English was a controversial subject in the seventies. At seventy most people have retired. To re-tire means "to put new tires on a vehicle." Some vehicles such as hovercraft have no wheels. Wheels go round (Enkvist, 1990, p. 12).

The text in this example has plenty of lexical cohesion (lexical repetition), but it is difficult to imagine any consistent plausible text world (Enkvist, 1990; Oller & Jonz, 1994).

By the same token, a text with missing or misused cohesive devices may still be seen as coherent and comprehensible through means other than cohesion (Koshik, 1999). The following example was composed by a nonnative speaker of English:

Someone come my house. Says give me money. Husband take gun shoot. Go outside die. Call police. Emergency 911. Policeman come. Take black man go hospital die. (Koshik, 1999, p. 11).

Koshik comments that this story has no grammatical cohesive devices and only instances of lexical cohesion (e.g., *someone/black man*, which are coreferential, and *police/policeman*, which contribute to topic continuity), but that the story is still seen "as a coherent whole" by readers who are native speakers of English (p. 12).

Means other than cohesive devices for establishing a sense of coherence include reader expectations of finding coherence and the frames provided by genre expectations, such as predictability inferred on setting and time sequence in narratives and the structure of ordinary conversation, such as the adjacency pair, question-answer and request-compliance (Koshik, 1999; Sacks, Schegloff, & Jefferson, 1974; Schegloff, 1990). For instance, in a conversation or a written story, even when the speaker or writer misuses a reference marker or a conjunction (e.g., *he* versus *she*, *one* versus *it*, *and* versus *but*; see text example in Appendix C, line 2 and example in Koshik, 1999, p. 14-15), the listener or reader may often have no problem with understanding the referent and a correct supra-sentential connection because the conversational structure and the setting and characters introduced in a narrative provide "a strong predictability even before the wrong form is used" (Koshik, 1999, p. 14).

Coherence is also established by the mutual interaction of the writer and reader to make sense of the text based on their shared background knowledge outside the text (Bamberg, 1984; Koshik, 1999; Renkema, 1993; Smith, 1984). Let us look at the following example used in Enkvist (1990):

The net bulged with the lightning shot. The referee blew his whistle and signaled. Smith had been offside. The two captains both muttered something. The goalkeeper signed for relief (Enkvist, 1990, p. 12).

This text has coherence, although it lacks overt grammatically describable cohesion markers such as repetition and reference markers (Enkvist, 1990; Oller & Jonz, 1994). The text becomes coherent when certain knowledge of the world, that is, knowledge of a soccer game in this case, is applied (Renkema, 1993, p. 35). Hence, a coherent text conforms to a consistent world picture for the reader, and therefore the meaning in such a text is summarizable, comprehensible, and interpretable (Oller & Jonz).

In summary, cohesion is a more grammatical, formal, and explicit property, whereas coherence is a matter of relevance, more pragmatic in nature, and a more global property (M. Celce-Murcia, personal communication, 1996). While cohesion is easily divisible into distinct subdimensions, coherence is not susceptible to subdimensions characterized by overt markers.

Content

Content is the semantic domain of language. In this study content is defined as the relevance of a written text to a given task, as well as thoroughness, persuasiveness, and creativity consistent with task expectations. The quality of content is thus viewed as the degree to which the writing impresses the reader in terms of these criteria.

Like coherence, content is not divisible on the basis of overt grammatical markers. The quality of content can be evaluated within a phrase or a sentence, but it can also be evaluated in a more global, holistic context such as many pages taken as a whole. An incoherent text with disjointed connections cannot communicate content effectively. For these reasons, content and coherence are thought to be closely related.

Grammar

Grammar refers to morphology and syntax and best represents the linguistic domain of language. Grammar is evaluated by the range of grammatical features and the extent of grammatical errors in the text. In evaluating the extent of grammatical errors, it is useful to classify grammar errors into critical errors and minor errors depending on the seriousness of the effect on reader communication. Critical errors are defined as errors that seriously impede communication, for example, a sentence-level structure and a syntactic chunk missing. These would have to do with grammar at a global level. Minor errors, on the other hand, are defined as errors at local levels such as incorrect or omitted morphemes, for example, third person agreement and tense agreement at the local level, which do not cause difficulties in comprehension.

Viewing grammar as a global text quality makes sense because learners' lack of grammar leads to a limitation on producing a text that has a quality and

length reasonable enough to get ideas across, while serious errors in grammar are likely to cause communication obstacles. Grammar is related to coherence and content in that grammar is also a global quality and because content and coherence do not exist in a vacuum but are represented in forms realized by grammatical structures. Grammar is also connected to cohesion in that grammar can also be a local text quality as indicated in the examples of minor errors (above) and in that cohesion is partly expressed through grammar (Halliday & Hasan, 1976/1993) and is a more formal and explicit property.

Text length

Text length defined in this study is the total number of words in a writing sample.

METHOD

In this section, the method is described in terms of study participants, study variables, the writing task, and administrative and scoring procedures.

Study Participants

A total of 192 students (97 first graders and 95 second graders) participated in the study. They were students enrolled in the Korean/English Two-Way Immersion Program and students in typical English-only classes. While this study does not address a cross-group comparison, it still may be informative to outline group characteristics to illustrate the diversity of learning contexts included in this study. Group characteristics are thus outlined in Table 2.

Table 2: Group Characteristics

Group		# of Subjects	Level of English Oral Proficiency	Curriculum Instruction in English
Immersion Program	Korean-Americans	66	100% LEP upon entering Kindergarten	30% (Kindergarten) to 50% (Grade 2)
	Non-Korean-Americans (EP**)	45	95% EP	
English-only classes	EP	81	100% EP	100 % (all grades)

* LEP: Limited English Proficient

** EP: English-Proficient

Each of the three groups shown in Table 2 had first graders and second graders. The curriculum difference between the immersion classes and English-only classes was the percentage of instruction conducted in English. Specifically, in the

immersion program, 30% (in Kindergarten) to 50% (in Grade 2) of instruction was conducted in English. In contrast, in regular English-only classes, 100% of school instruction was in English for all grades. The following paragraphs will give further information about the subject characteristics of the groups.

Immersion groups

Three elementary schools in the Los Angeles Unified School District (LAUSD) participate in the immersion program, and each school includes one immersion class per grade level. The immersion groups of this study consisted of the three first-grade classes and the three second-grade classes that come from these schools.

The Korean-American students in the immersion program: This group consisted of Korean-American students whose home language is Korean. Upon entering kindergarten, they were identified as *Limited English Proficient* (LEP).¹ This identification was based on the district-administered test of oral proficiency in English, called the pre-Language Assessment System (pre-LAS), which was designed to identify an initial oral proficiency level at the time of entering schools.² Although the Korean-American students were identified initially as LEPs, it is noted in Bae (1997) that at the end of the second grade, the English writing skills were on par with their peers in typical English-only classes.

English-dominant students in the immersion program: Non-Korean-American students are from Euro-American, Hispanic, Tagalog, Chinese, or Japanese background. 58% of the students in this group used English-only at home (EO students), 42% of this group used both English and another language (Spanish, Tagalog, Japanese, or Korean) at home with the exception of two students who had only Chinese or Spanish as a home language. All of these non-Korean-American students (except the latter two students, who were LEPs) were classified as English-Proficient (EP) upon entering Kindergarten, according to the LAUSD's language classification criteria (see note 2).

English-only classes

The students from English-only classes came from two schools where the immersion program is operating concurrently. The two schools had middle to upper level status with regard to general academic achievement based on the results of a national test, the Comprehensive Tests of Basic Skills, or CTBS (LAUSD Information Technology Division, 1996).³ The English-only classes from these schools consisted of two first-grade classes and two second-grade classes. These classes are typical English-only classes in that they receive regular curriculum instruction common to English-medium classes at LAUSD, and, like other English-only classes at LAUSD, these students are English-Proficient students in terms of an English oral proficiency level. In essence, all of the students from English-only classes were English-Proficient students in terms of their English

oral proficiency based on the district's classification criteria (A. Shoji, M. Hicks, & R. Rudnick, personal communication, July, October, 1996).

Study Variables

The theoretical definitions were articulated in the previous section "Construct Definitions." The measurement of these constructs was operationalized into observed variables by implementing a writing task and the scoring criteria to be described below. The observed variables produced in this process are coherence, content, grammar, text length, and the five subconstructs of cohesion, which includes reference, lexical ties, conjunction, ellipsis, and substitution.

Writing Task

Writing prompt

All subjects were given a series of pictures and were instructed to make up and write a story based on the picture series (see the instructions in Appendix A). The picture series is given in Figure 1.

Figure 1*



* Story and illustrations by Jungok Bae and Hyesug Lee.

Genre of the writing test

The writing prompt was intended to elicit narrative, which was considered useful in this assessment for the following reasons. Narrative is a socially and academically valued skill, and children are often called upon to read and tell stories at home and in school to improve reading and writing skill development (Peterson & Dodsworth, 1991). Narration is thus a common experience to children, and the ability to narrate develops early in childhood. In particular, a picture-based narrative task with connected scenes was considered relevant since the scene connection provides a useful means to examine language skills beyond the

sentence level (Ripich & Griffith, 1990). Through the pictures, both the writer and the reader share maximum background knowledge consistent with what is being written, thus minimizing possible comprehension obstacles between the writer and the reader. All in all, the visually connected set of pictures provides a common contextual ground for comparing narrative production across the subjects (Berman & Slobin, 1994, pp. 41-42).

Administrative Procedure

To promote consistent test administration, written and spoken guidelines for delivering instructions for test administration were given to the teachers on the day of the testing at each separate class-by-class test administration. The teacher-delivered instructions were intended to give students their own teacher's language and delivery in a style familiar to them, thus creating a comfortable testing environment. To assure consistent test administration across the classes, the same test coordinator (the author) was present at all class test administrations. The test was given toward the end of the school year, in late May through July, 1996. The time allotted to the actual writing was up to 30 minutes, with most students finishing the story-writing in 20 to 25 minutes, while some students required the maximum time.

Scoring

Coherence, content, and grammar

Dimensions such as coherence and content do not have overt linguistic markers that are countable. Grammar shows overt linguistic markers, but the range of grammatical features is countless. Therefore, a holistic judgment based on the rating scale was made in scoring these three dimensions of language. In other words, the scoring uses a componential or analytic scoring method with a holistic judgment made separately for coherence, content, and grammar within each component. Each dimension was scored independently by two graduate students from the Department of Applied Linguistics and TESL at the University of California, Los Angeles who were native speakers of English. Their scores were averaged to be the score for each individual. For ratings that showed discrepancies with more than one scale point, a third rater assigned a rating: The closest two of the three ratings were averaged to be the score for each individual.⁴ All samples were shuffled together before rating, and students were identified only by their identification numbers to prevent raters' possible bias concerning any ethnic group and school grade.

Rating scale

The following generic scale (adapted from Bachman, 1989) formed the basis of the rating scale for scoring coherence, content, and grammar:

0	1	2	3	4
Zero	Limited	Moderate	Extensive	Complete

The characteristics for the five scale points were specified for each component of language. The two ends of the above scale were defined to provide an absolute scale (Bachman, 1989, pp. 251-258; 1990 pp. 340-348) as follows: The one end point (0) represented zero or very little ability; the other end point (4) represented a complete level of the written English language ability for second graders. The second graders were the highest grade used in this study. Thus, the characteristics for determining the scale point of 4 for each of the language components, including benchmark samples, were the ideal level of language use observed in the best second graders' written samples. In addition, ratings with a 0.5 decimal point for each scale point (that is, 0.5, 1.5, 2.5, and 3.5) were incorporated (see Bae, 2000, p. 82, for advantages in allowing decimal points). Expectations relevant to these student levels were also considered in forming the scale descriptions (for example, colloquial expressions, childlike expressions, and errors in spelling were considered acceptable at this level). These characteristics thus formed the criteria. The second graders under consideration consisted of the groups from the immersion program and the English-only classes, which included native and nonnative speakers of English at grades 1 and 2. (See the scoring criteria and selected best sample in the Appendix.) The absolute scale was intended as a common metric scale for all subjects in this study. This common scale is used to make the results of the assessment apply not only to the immersion students but also to the students from English-only classes.

Cohesion

In contrast with coherence, content, and grammar, cohesion has explicit linguistic markers that are countable; thus, counting the number of markers was considered a method that would give a more accurate account of the dimensions of cohesion demonstrated in the writing samples. Thus, appropriately used cohesive markers were counted in each of the following areas of cohesion: reference, lexical ties, conjunction, ellipsis, and substitution (See Table 1 for types of cohesion with examples). Two raters counted the number of these cohesive markers independently, excluding errors, for the randomly selected one-third of the entire samples. Since counting the frequency of something was an objective procedure, compared to judgment against the rating scales, after ensuring the high interrater agreement (over .950, see Table 4), one rater counted the number of cohesive markers for the rest of the samples.

Analysis

Descriptive statistics, interrater reliability estimates, correlations, and multiple regression were calculated on SPSS release 9.0 and SAS release 6.11. Qualitative, descriptive analyses were conducted independently of ratings.

RESULTS

Descriptive Statistics

Table 3 reports descriptive statistics for the scores for each variable. The entire data as a single group ($N = 192$) was used for this purpose.

Table 3: Descriptive Statistics

	Measurement unit	N	Mean	S.D.	Skewness	Kurtosis
Grammar	0 to 4 Rating Scale	192	2.61	1.00	-0.88	0.24
Content		192	2.48	1.04	-0.71	0.02
Coherence		192	2.20	1.11	-0.28	-0.69
Cohesion						
Reference	Frequency of Correct Usage	192	14.29	9.65	1.68	6.11
Lexical ties		192	24.72	13.48	0.98	2.13
Conjunction		192	5.18	3.88	1.22	2.54
Ellipsis		192	0.31	0.75	3.85	20.35
Substitution		192	0.10	0.34	3.40	11.79
Text length	Number of Words	192	67.51	35.53	1.20	3.81

As shown in Table 3, all variables, except for ellipsis and substitution, showed skewness and kurtosis within or slightly greater than ± 2 , thus showing a normal or approximately normal distribution of scores for these variables. (Skewness greater than ± 2 indicates that the tail of the distribution is to the right or left compared to a bell curve or a normal distribution; kurtosis indicates that a distribution is either peaked or flat compared to a normal distribution.) The examination of a bar graph for each variable (not provided due to space limitation) confirmed the approximately normal distributions for these variables.

Looking at the means taking all students together, coherence, content, and grammar showed a mean score of around 2.5 on the 0-to-4 scale. The average number of words used in the essays was about 67; the average of number of occurrences of reference markers was 14, and the average number of occurrences of conjunction markers was around 5.

In contrast, ellipsis and substitution were found to have far fewer occurrences in the compositions. The average occurrence in the essays was less than 1 (with a mean of 0.3 for ellipsis and 0.1 for substitution). Ellipsis and substitution showed a deviation from a normal curve as indicated by a very small degree of dispersion of frequencies of occurrence (near zero SD's: 0.75 and 0.34). These

two variables showed a somewhat positively skewed curve (with a skewness of 3.85 and 3.40, respectively), indicating that the frequencies of occurrence clustered around the fewer occurrences. These variables also showed high kurtosis (20.35 and 11.79 each), indicating a sharp, peaked curve around the low frequencies of occurrence. These results indicate that, compared to other components that appeared extensively in the essays, in the vast majority of essays ellipsis and substitution had extremely few occurrences.

Rater Correlation and Reliability

Since two raters independently assigned a rating for each of the components of coherence, content, and grammar, rater agreement and the consistency of ratings are of concern for these components. Rater agreement is represented by the interrater correlation, and consistency (reliability) of two ratings by rater reliability. Thus interrater correlation coefficients and rater reliability coefficients were estimated for the entire data for each variable (See Table 4).

Table 4: Rater Correlations and Rater Reliability

		Interrater correlation		Alpha interrater reliability
		Pearson	Spearman	
Coherence		.907	.904	.951
Grammar		.881	.830	.937
Content		.920	.899	.958
Cohesion:	Reference	.999		.980
	Lexical	.998		.993
	Conjunction	.997		.998
	Ellipsis	.994		.997
	Substitution	1.000 *		1.000*

Note

*The perfect rater reliability indices for substitution were due to the easily noticeable nature of substitution markers (e.g., *So he did.*) and very rare frequencies of substitution (only a total of 5 occurrences from the randomly selected 80 samples), which resulted in the easy agreement of two ratings, and in addition, discussions during the rating session.

As an index of rater agreement, Pearson correlation coefficients were calculated for the components coherence, content, and grammar ($N = 192$). The Pearson r indicates the index of the strength of the linear relationship between two variables (in this case, two ratings independently assigned by two raters). This coeffi-

cient (r) is appropriate for interval or continuous data and assumes normality of data. The variables for these three components become continuous data by allowing a decimal point 0.5 on the rating scale (see Bae, 2000, pp. 81-82 for a detailed explanation), and normality is largely met. Therefore, the Pearson r can be used as a proper measure of rater agreement for these components. Pearson r ranged from .881 to .920. As a supplement, the Spearman's rank r is reported. This index provides another measure of the linear relationship between two variables (two ratings) for ordinal or interval data that do not satisfy the normality assumption. Spearman's coefficients ranged from .830 to .904.

As an index of rater reliability, alpha interrater reliability coefficients were calculated. The alpha (α) reliability coefficient reported in this table represents the internal consistency of two scores for each variable, and it is based on the average covariance between the two ratings or scores on the variable. Alpha ranged from .937 to .958 for these components. All coefficients thus indicated a highly acceptable degree of rater agreement and reliability or consistency of two ratings for each component.

In addition, Pearson coefficients and alpha were used as a supplement for the cohesion variables for the randomly selected samples ($N = 80$). As indicated previously (see Scoring section), counting markers of cohesion was a highly objective procedure compared to judgment against rating criteria. As expected, the indices showed near perfect rater agreements and reliabilities (over .980).

Correlations

Understanding a focal construct in relation to another is essential in construct explication. A statistical index to indicate the relationship between two variables is a correlation coefficient (r). Since the present study is interested in the extent to which a particular construct is related to another, correlations will be reported and referenced in the discussions of the constructs in the subsequent sections.

Table 5 reports correlations for the variables of coherence, content, and grammar, the five dimensions of cohesion, and text length. Pearson coefficients were used because the variables were based largely on the continuity of scores. The scores of coherence, content, and grammar were summed up, and will be referred to as the "overall" writing quality throughout the paper.

The Constructs

In this section, the results of analyzing the compositions are provided and discussed for each construct. Text length will be discussed first. Cohesion will be discussed with an elaboration made in the sub-dimensions. Subsequently, coherence, content, and grammar will be discussed within a section on overall writing quality.

Table 5: Correlation Matrix

	Grammar	Content	Coherence	Reference	Lexical Ties	Conjunction	Ellipsis	Substitution	Length
Grammar	--								
Content	.797	--							
Coherence	.797	.900	--						
Reference	.650	.750	.766	--					
Lexical	.637	.769	.767	.901	--				
Conjunction	.364	.472	.413	.484	.534	--			
Ellipsis	.303	.343	.372	.567	.414	.108	--		
Substitution	.155	.225	.245	.327	.343	.273	.099	--	
Length	.650	.762	.760	.927	.937	.636	.455	.388	--
Overall *	.914	.955	.957	.768	.771	.442	.361	.223	.770

Note

N = 192 for all variables. All correlation coefficients above were significant ($p < .001$) except .099 for the ellipsis-substitution relationship.

* Overall = Sum of Grammar, Content, and Coherence.

Text length

The length of a writing sample is defined as the total number of words. Length was considered part of fluency, and its relations to other components of abilities were examined. Raters were asked not to look at length in judging coherence, content, and grammar once an essay exceeded a certain threshold-level length (in this study, approximately 30 words out of a range 7 to 247 words), although an essay that was too short was considered limited in coherence, content, and grammar unless the writing was error-free and persuasive in content. From Table 5, we can see that text length was highly correlated with coherence ($r = .760$), content ($r = .762$), and grammar ($r = .650$). This result supports the notion that in general a fluent writer can write a longer essay within acceptable overall qualities. Text length was also highly correlated with reference ($r = .927$) and lexical ties ($r = .937$)—we will discuss this point in the subsequent section on reference and lexical ties. However, text length showed moderate and small degrees of relationships with ellipsis ($r = .455$) and substitution ($r = .388$), indicating that a longer essay does not necessarily exhibit more occurrences of ellipsis and substitution markers.

Dimensions of cohesion

Let us examine the subdomains of cohesion. The frequencies of occurrence of cohesive markers were counted separately for each domain of cohesion, together with the relative percentages, for the entire sample as a whole (see Table 6).

Table 6: Frequencies of Cohesive Devices (for the Entire Data Set)

<i>Dimensions of Cohesion</i>	<i>Mean Occurrence</i>	<i>Percentage of the Total Occurrences</i>
Lexical Cohesion	24.7	55.6%
Reference	14.3	31.8%
Conjunction	5.2	11.7%
Ellipsis	0.3	0.6%
Substitution	0.1	0.2%
Total		100%

As Table 6 shows, lexical cohesion was the dominant pattern of cohesion observed in the student narratives (55.6% of total occurrences of all cohesive markers), followed by reference (31.8% of total occurrences) and conjunction (11.7%). Instances of ellipsis and substitution occurred relatively less frequently in the story-writing (each type less than 1%).

In the following sections the subdomains of cohesion are discussed in detail.

Reference and lexical ties

Reference: The types of reference and their relative frequencies observed in the writing samples are given in Table 7. The dominant reference type was pronominal forms (58.8% of total occurrences of references), followed by proper nouns (22.8%). The definite article *the*, demonstratives, and comparative reference occurred relatively less frequently (15.1% to 0.5%). The prominent use of pronouns and proper nouns appears to be due to their role as head nouns, primary information for reference, whereas definite articles and demonstratives are modifiers.

Table 7: Types of Reference and Relative Percentage of Occurrences

<i>Types of Reference</i>	<i>Percentages</i>
Pronominals: (<i>he/she/him/their</i>)	58.8%
Proper nouns: (<i>Billy, Jack's</i>)	22.8%
Definite article: <i>the</i>	15.1%
Demonstratives: (<i>this/these, that/those, here, there</i>)	2.8%
Comparatives (<i>bigger, the same, both</i>)	0.5%
Total	100%

Lexical ties: This study echoes the observation that among areas of cohesion lexical cohesion is the most unexamined area since the studies by Halliday and Hasan (1976/1993, 1989). As such, identifying words that contributed to lexical ties was not as clear-cut as analysis of other areas of cohesion; the same is true with the quantification of lexical ties. Thus, there is still a need for future research to refine theoretical definitions and the domain of lexical cohesion. Nonetheless, in the current study, lexical ties were calculated by counting the number of the words in the essays that belonged to any of the domains of lexical cohesion, such as synonyms, antonyms, collocations, part-whole relations, general-specific lexical ties (see Table 1 for examples), and a stream of main ideas represented by main verbs. Phrasal verbs and idiomatic expressions (e.g., *go home, put up, come in*) were counted as single lexical items.

Reference and lexical cohesion: As seen in Table 5, use of reference and lexical ties were highly correlated ($r = .901$). The use of reference and lexical ties was moderately or highly correlated with other writing qualities such as grammar ($r = .650, .637$), content ($r = .750, .769$), and coherence ($r = .766, .767$). These high-to-moderate correlations suggest that acquisition of reference markers and vocabulary is critical to extending an essay and enhancing one's overall writing quality.

Another notable observation is that reference and lexically-tied words were highly correlated with length ($r = .927$ with reference; $r = .937$ with lexical ties). This correlation is not surprising because the longer a writer composes, the more reference markers and lexical ties the writer will naturally use.

A note is also necessary about the high degree of correlation between these two types of cohesion and length. The greater number of cohesive markers could be explained partly by the longer writing samples. The longer length and the greater number of cohesive markers used could be interpreted as an indicator of fluency. However, caution should be taken not to infer that students who demonstrated a smaller number of cohesive markers in short texts would not show competency in utilizing cohesive markers in their longer writing samples.

Conjunction

For the purpose of this study, the definition of *conjunction* as a form of cohesive tie is confined to coordinating conjunctions between two independent main clauses (e.g., *The boy saw a girl, and she was crying*). Subordinating conjunctions connecting a main clause and a dependent clause (e.g., *I saw a girl when I was walking*) are not considered cohesive conjunctions; thus they were not counted. Halliday & Hasan (1976/1993) introduce four types of conjunctive relations: *additive*, *adversative*, *causal*, and *temporal* (see Halliday & Hasan, pp. 238-273 for detailed examples of the words and phrases that express these meanings). These four types of conjunctions and their frequency of use in the children's writing were examined and are reported in Table 8, taking all writing samples together.

Table 8: Types of Conjunctions and Their Relative Percentages of Occurrences (# of total occurrences: 1141)

Additive		Adversative		Causal		Temporal	
and	23.0%	and	0.5%	and	3.9%	and	30.1%
but	1.1%	but	1.8%	so	12.4%	then	15.5%
				other *	0.3%	and then	7.4%
						adverbial	3.9%
Total (100%)	24.1%		2.3%		16.6%		56.9%

* Ordinarily, subordinating conjunctions are not treated as conjunctions in the form of cohesive ties. However, some students in this study used a subordinating conjunction as a reply to a question in a question-answer sequence: e.g., " Jill went to her, Why are you crying? *Because my umbrella broke*." (from a first grader's sample). These subordinating conjunctions were counted as conjunctions in the form of cohesive ties because they occurred in an independent sentence, creating a connection across sentences, which represented speaker turns in the narrative.

Nearly 57% of all occurrences of conjunction were temporal. The next most frequent occurrences were additive (24.1%), followed by causal (16.6%) and adversative (2.3%). Another noteworthy feature of conjunction use at these grade levels was the extensive use of *and*, which indicates that the coordinating conjunction *and* is an early acquired conjunction, used for multiple functions:

He saw a girl and her umbrella was broken. (Additive)

The girl is getting wet and the boy is not wet. (Adversative/Contrastive)

My umbrella is torn and I'm soaking wet. (Causal)

They were drawing and Amy's mother came with food. (Temporal)

The children also often used *and* as an "utterance initial filler to indicate more is to come" (Berman & Slobin, 1994, p. 176) as in:

Paul was walking . . . And it was raining . . . And Paul saw Tiffany. And Paul said, come in.⁵ lets walk home. I'll go with you. OK. And they were walking and walking . . . And then finenuly they went to Tiffany house. And Tiffany said, come in . . . And Paul came in and she was happy . . .

It is noted that the use of conjunctions was only moderately related to the overall writing quality, such as grammar ($r = .364$), content ($r = .472$), and coherence ($r = .413$), which means that more frequent use of coordinating conjunctions does not necessarily contribute to the overall writing quality. This was true because many of the conjunctions used in these children's essays, particularly *and* and *then*, were non-essential elements, although not inappropriate. There were 11 students who used few or no coordinating conjunctions but received a high rating (moderate to complete) in coherence and content. On the other hand, one student connected all sentences with a coordinating conjunction, *and*, producing an entire writing sample of 108 words in one sentence, but still exhibiting high qualities of coherence, content, and grammar. In general, second graders' samples demonstrated a more diverse use of coordinating conjunctions beyond *and* to also include frequent use of *but*, *so*, *then*, *and then* and, less frequently, common temporal adverbials such as *first*, *last*, *after that*, *finally*, and *soon*.

Ellipsis and substitution

Typically, ellipsis is known to occur in responses in spontaneous conversations but is seldom used in formal writing. As such, ellipsis had far fewer occurrences than lexical ties and reference (see descriptive statistics presented in Table 3). However, children frequently introduced dialogues and interactional conversational expressions into their story progression, which can be interpreted as young writers' individual choice of rhetorical styles that enriched the narratives. Use of ellipsis was observed in such dialogues as in "You want to draw? Sure, Timmy said." and the writing of a few adept students:

- (1) "Thomas and April ate some of the fruit but not all"
 (2) "Where are you going Cindy?" "To my house!" ... So they went inside [She] put [her] backpack away in her room. Even Billy too!"

Like ellipsis, substitution is also a speaker/writer choice and not a compulsory feature (McCarthy, 1991, p. 43). Accordingly, substitution seldom occurred in these writings, except that several students demonstrated an elegant use of substitutions as in:

- (3) "Dennis went up to her and said, "Let's share my umbrella." So they did."
 (4) "Paul's friend is happy and so is he."

Similar to conjunction, ellipsis and substitution showed weak relationships with overall writing quality ($r = .361$, ellipsis and overall; $r = .223$, substitution and overall; see Table 5).

Errors in cohesion

Errors in the use of cohesive ties in each sample were marked by two native speakers of English. Table 9 summarizes the patterns of those errors across the entire data set, with examples provided.

As shown in Table 9, the majority of errors involved problems with reference: unclear references (26.5%) and misuse of *a* or *the* and omission of determiners needed for reference (56.4%). The rest of the errors involved conjunctions and minor grammatical and syntactic errors (e.g., in shes house) at a local level. Despite such errors, the general meaning was inferrable in the text and from the shared contextual schemata provided by the picture series.

A note is necessary about the comparison made between the English-language background and the Korean-language background groups in examining errors in cohesion. Unlike coherence and content, errors in cohesion are easily definable, noticeable, and countable; therefore, subjectivity of cross-group comparison is not a problem with cohesion errors. Compared to the vast range of patterns of errors in grammar, patterns of errors in cohesion are easy to categorize, and the range is not unlimited. Although this paper does not intend to make cross-group comparisons, motivated by its relevance to the understanding of the focal construct, dominant error patterns in cohesion were compared between the two language groups. Dominant error patterns differed depending on the language background. For the English-Proficient (EP) students from both immersion and English-only classes, the dominant error pattern was unclear references: 33% of the immersion EP students' total error occurrences and 25% of total error occurrences made by students in English-only classes were unclear reference. In contrast, the dominant pattern of errors for the Korean-American (KA) students involved misuse of either *a* or *the* (about 59% of KA's total errors) and omission of determiners needed for reference (19% of KA's total errors). This was apparently due to the

Table 9: Errors in Cohesive Markers (Total Number of Occurrences: 204)

<i>Patterns of Errors</i>	<i>Percentages</i>	<i>Examples</i>
Unclear references	26.5%	Sudden switch of references; e.g., from third person pronouns to a first person pronoun; exophoric, unclear use, and wrong use, which are inferable or incomprehensible.
Misuse of 'the'	24.0%	<u>The</u> Dennis was helping the Helen. The boy went to <u>the</u> her home. Once there was <u>the</u> boy.
Misuse of 'a'	16.7%	Mom brout a food and a fruits. Drew a boats. Drew a ship on a paper.
Omission of determiners that make references (a/the/pronominals)	15.7%	__ Girl waved to __ boy. They went to __ boy's house. The picture was __ ship.
Unnatural use of conjunctions	10.8%	A boy named Eddy had a umbrella at school. <u>But</u> when he was going home he saw a girl.
Other	6.4%	Came over <u>shes</u> house. I shared <u>my's</u> with her. <u>She</u> mom gives...
Total	100%	

transfer from the Korean language, which allows null articles in the places where articles are obligatory in English.

Overall writing quality

Let us turn to the overall writing quality, which includes coherence, content, and grammar.

Coherence

Earlier, we discussed potential factors for establishing coherence: (a) shared background knowledge of the world between the writer/reader, (b) cohesive markers, and (c) the frames underlying genre expectations. Results concerning these factors will be discussed below.

First, the pictures were sufficient for providing writers and readers the shared contextual schemata to facilitate comprehension; thus, the shared background knowledge necessary for a coherent text was well pre-established.

Second, with reference to measuring cohesion as a factor for establishing coherence, a question that can be asked is: How much does the use of cohesive devices account for, or contribute to, coherence? To investigate this question, multiple linear regression analysis was used with coherence as the dependent variable and the five types of cohesion as multiple independent variables or predictors. The entire data as a single group ($N = 192$) was used for this purpose. Multiple regression is a statistical procedure for analyzing the collective and separate effects of multiple independent variables on the dependent variable. With this method, we can determine which of the independent variables best predicts or accounts for the dependent variable (Hatch & Lazaraton, 1991; Pedhazur, 1982). The results are summarized below.

**Table 10: Summary of Multiple Regression: Coherence
Accounted for by Cohesion Variables**

<i>Significant Predictors</i>	<i>Simple r with Coherence</i>	<i>Cumulative R^2</i>	<i>Changes in R^2</i>	<i>Beta</i>	<i>Significance (p)</i>
Lexical ties	.766	.588	.588	.406	.000
Reference	.767	.618	.030	.400	.000

Referential and lexical ties were found to be substantively significant ($p = .000$) predictors of coherence. These two types of cohesion, lexical ties and reference, collectively, accounted for 61.8% of the total variance in coherence ($R^2 = .618$). On the other hand, the other types of cohesion (conjunction, ellipsis, and substitution) were not significant predictors ($p = .944, .517, .555$, respectively) and did not enter the final regression equation.

The magnitudes of the effects of the two significant predictors were indicated by the standardized regression coefficient, beta. The beta coefficients indicated that lexical cohesion and reference showed nearly the same magnitudes of strength as a predictor of coherence (beta = .406 for lexical; beta = .400 for reference). At the same time, these two variables showed a large proportion of shared variance in predicting coherence, as indicated by their high correlation and the small change in R^2 in the regression.⁶

Thus, for this writing task, it is concluded that the degrees to which the use of cohesive devices account for, or contribute to, coherence vary from highly significant (lexical and reference) to very little (conjunction, ellipsis, and substitution) depending on the subdomains of cohesion analyzed.

Third, it was pointed out that coherence has to do with a plot-motivated overall structure and the frames underlying genre expectations, such as setting, topic continuity, and a coda in narratives. As such, the following attributes characterized a high rating for coherence in this study: a clear presence of elements for

introductory remarks, a coda (conclusion), and elaborated connections filled in at every important local point; elaborated connections between ideas; a comprehensible and consistent idea stream throughout the essay. As indicators of this, the study calculated the percentages of students who wrote an explicit introduction and a coda in the writing samples and compared the grade performance. The results showed that the vast majority of second graders (90.5%) and the majority of first graders (63.6%) used a clear introduction. These results suggested that they possessed a concept of how typical story books begin. Most of those who had an explicit introduction used either classical opening words such as *once upon a time* and *one day* or some form of temporal or locative adverbials to set up the stage and the time (e.g., *one evening, on the street, yesterday . . . from school*).

The ability to close the writing with an explicit ending was relatively less developed, suggesting that an awareness of conclusions is a cognitively higher-order or later developed skill than that of awareness of introductions. Only 40.8% of second graders and 14% of first graders drew an explicit conclusion from what was given at the final scene. Examples of the presence of a coda are illustrated below (from second graders):

(5) "... and I thought the girl and I are best friends now."

(6) "... And from that very last day the children would play together Nice things together."

Content

The rating criteria in this study specified that a high rating for content was determined by persuasiveness and creativity within task relevance and thoroughness of the content with respect to the picture prompts. Content was analyzed by examining the essays in terms of the following two categories: (a) mere descriptions of what is visually given in the pictures, and (b) something beyond the visually given content, such as interpretations, evaluations, personal feelings, or imagination that enriched the content within task relevance. Such enriched content in these children's writing is illustrated below from different samples. (The spelling errors in the samples below were not judged in any areas of scoring):

(7) "A girl was walking with a broken umbrella and she was sad so the boy share his umbrella with her they where happy."

(8) "They remembered they had a homework. They took out their homework. It was drawing a boat that rowed around the sea."

(9) "I felt sorry for her so I shared my umbrella with her . . . I asked if she wanted to draw pictures . . . My mom thought we should hang the picture's on the kitchen wall."

(10) "Mother said I like you're activities I'll think I'll paste them on the wall."

(11) "So Jill and Sue had a lought of fun And there best friend."

(12) "Amanda and Eric drank all the milk because they were so hungry. And it was a feast."

In the previous section, the presence of an explicit introduction and ending in the stories was compared by grade level, providing a useful means to examine the development of coherence. In examining the development of content, analysis by grade level was also useful. For each grade level, a percentage of students who went beyond mere descriptions of what is visually given in the pictures (that is, category [2] above) was calculated. 34.3% of the first-grade compositions and 72.7% of the second-grade compositions belong to this category. Thus second graders demonstrated a far greater degree of ability to use diverse expressions, make inferences, and make relevant associations about what was not explicitly provided in the pictures than did the first graders. The development of content, like that of coherence, seems to be related to the cognitive maturational development in young children.

Content and coherence: Whether content and coherence are actually qualitatively the same ability traits is a separate question. However, content and coherence showed an extremely high linear relationship with each other ($r = .900$, see Table 5). The two constructs seem to be very closely associated with each other.

Grammar

Relationship with coherence/content: As seen in Table 5, grammar showed a high correlation with coherence (.797) and content (.797). Thus grammar could be treated as a global writing quality factor, together with coherence and content. Viewing grammar as a global writing quality is reasonable because without adequate competency in grammar it is unlikely that learners can produce writing with quality and text length reasonable enough to communicate ideas. Grammar errors were defined as critical errors or minor errors depending on the seriousness of the effect on reader understanding (See "Construct Definitions"). Writing samples with critical errors received low ratings. Minor errors, on the other hand, were considered tolerable enough to allow the writer to get moderate to high ratings.

Relationship with cohesion: Among areas of cohesion, grammar was most highly correlated with reference and lexical ties ($r = .650$ and $.637$ respectively) and it was moderately or very weakly associated with conjunction ($r = .364$), ellipsis ($r = .303$), and substitution ($r = .155$) (See Table 5).

CONCLUSIONS

In this study, cohesion, coherence, content, and grammar in English writing samples composed by 192 first or second graders were investigated within the context of immersion and English-medium classes. Following are the conclusions to the research questions concerning (a) the characteristics of cohesion and coherence and the interrelations among cohesion, coherence, content, grammar, and text length, (b) prominent types of cohesion, and (c) errors in cohesion.

Characteristics of Cohesion and Coherence and the Interrelations among Cohesion, Coherence, Content, Grammar, and Text Length

As shown in Table 5, the correlations showed that coherence and content were highly related ($r = .900$). Grammar showed a high correlation with coherence (.797) and content (.797). Among subdomains of cohesion, the use of reference and the use of lexical ties were highly correlated with each other, and they both were highly related with the length of a text. The ratings of coherence, content, and grammar were summed up, and were referred to as the "overall" writing quality. Referential and lexical cohesion showed relatively high correlations with the overall writing quality ($r = .768, .771$, respectively). However, ellipsis and substitution showed relatively weak correlations with the overall quality ($r = .361, .223$, respectively). This supports the idea that the subdomains of cohesion, while they play a distinct role within cohesion, are more local-level links than the global quality represented by coherence, content, and grammar.

Multiple regression analysis examined the effects of the cohesion variables on coherence (See Table 10). The results showed that cohesion differed in degree as a contributor to coherence depending on the subdomains of cohesion. Lexical cohesion and referential cohesion were found to be significant ($p = .000$) predictors of coherence. The two variables showed almost the same magnitudes of the effect on coherence ($\beta = .406, .400$, respectively). They both, collectively, explained approximately 61.8% of coherence, indicating the importance of the acquisition of reference markers and vocabulary as a factor to establish coherence. On the other hand, other types of cohesion (conjunction, ellipsis, substitution) were not significant predictors of coherence.

Awareness of an introduction in narrative writing was well-developed for these grades (grades 1 and 2); however, an awareness of conclusions was less developed at this stage.

Prominent Types of Cohesion

The most prominent types of cohesion observed in the narratives across the children in this study were lexical and referential ties (respectively 56% and 32% of total occurrences of all cohesive markers; see Table 6). Coordinating conjunction, ellipsis, and substitution occurred less frequently in the written narratives. This result, together with the interpretations made above, suggests that reference and lexical ties are more crucial and necessary while the other types of cohesive markers can be present or absent depending on writer/speaker choice. Meanwhile, dominant reference types (see Table 7) were pronominal forms (about 59% of total occurrences of reference) and proper nouns (about 23%). Prominent types of conjunctive relations (see Table 8) were temporal (nearly 57% of all occurrences of coordinating conjunctions) and additive (24%).

Errors in Cohesion

The dominant patterns of developmental errors in cohesion involved inaccurate reference across the grades and the language background groups: unclear

references and misuse or omission of *a/the* and determiners that make referential ties. The most prominent errors for English Proficient (EP) students (that is, English-dominant students) in both immersion and regular English-only classes were unclear references. However, a dominant error pattern for the Korean-American students in the immersion program involved misuse or omission of *a* and *the*, reflecting transfer from their first language, Korean, which typically does not use articles. All of these errors, in general, did not hinder comprehension of meaning in the context of the texts and in the presence of shared schemata provided by the pictures.

This study conducted a theoretically informed empirical investigation of cohesion, coherence, content, grammar, and text length in children's written narratives in English within the contexts of a two-way immersion program and English-medium classes. The data collected from the performance-based story-writing task via group testing and the descriptive and quantitative data analyses shed light on the global interrelationships among these constructs and their dominant characteristics. The substantiated findings provide evidence for the characteristics of these constructs that would not be provided by purely theoretical write-ups and typical case study analyses. At the same time, the constructs established in this study may well provide a basis for subsequent studies (see below). Instruments comprising the writing prompt and the scoring method can be useful in applications to other contexts and students with different linguistic backgrounds.

Limitations of the Study

The schools, classes, and teachers in this study serve as convenient samples; they were not randomly selected. Even when schools and classes are randomly selected, the number of schools and classes is often so small that we cannot say random sampling gives a representative group (Bentler, 1997; Hatch & Lazaraton, 1991). At the same time, the subjects in this study are elementary school children, and the linguistic backgrounds of these students consist mostly of English, Korean, and Spanish languages. The data based on several classes in this study provides a much stronger basis for generalizing findings than data based on case studies that typically involve several students. Considering the above-mentioned limitations, however, caution should still be taken in generalizing the results of the present study.

Implications for Future Research

Several implications and suggestions for future research are made with respect to potential contributions of this study. The first suggestion involves a more sophisticated analysis with latent variables (as opposed to ordinary observed variables). A latent variable modeling approach is used to design and test models of relationships that include latent variables, free of error of measurement. With this approach, more sophisticated relationships can be specified and tested as they may be illustrated as follows.

1. *Constructs used as latent variables:* The variables cohesion, coherence, content, grammar, and text length used in the present study are observed variables, which contain inherent errors in measurement. One of the advantages of a latent variable approach is its capability to decompose observed variables into latent variables free of error of measurement. Inferences from data thus become more refined at the level of latent variables (also called *factors*), controlling for measurement error. Thus, in a future study, cohesion and coherence could be investigated at the level of latent variables.

2. *Directionality:* The correlations used in the present paper give fundamental information about the degrees of relationships between variables, but a correlation does not indicate causal relationship or explanation.⁷ With a latent variable modeling approach, the directionality of influence among variables, if not causal relationships, can be specified. To use this approach fruitfully, however, the researcher should have a well-developed a priori theory to test. For instance, an excellent topic where directionality of influence can be specified and tested with this approach would be the contributions of subdomains of cohesion to coherence. Another example would be to use groups (background characteristics) as predictors of the constructs (see below).

3. *Factorial evidence for construct distinctiveness:* A latent variable modeling approach is well suited to perform hypothesis testing, a typical format of construct validation. One important topic with respect to construct validation that is not addressed by the present paper is to test for psychometric evidence for the distinctiveness of the specified constructs. Inquiries that can be addressed in this direction would include questions, such as: Are cohesion and coherence (and the subconstructs of cohesion) statistically found to be distinguishable factors, and are they found to be distinguishable from grammar and content at the psychometric level? A future study could investigate this topic using confirmatory methods that allow hypothesis testing within a latent variable modeling approach. To my knowledge, no studies have provided psychometric evidence of their distinctiveness (see, however, Bae, 2000 for the evidence for distinctiveness of grammar, content, spelling, and text length).

4. *Group comparisons:* Another important topic is the comparison of writing performance on these constructs between immersion and English-medium classes. Group comparisons inform program evaluation in achieving goals with respect to the achievement of students' English writing skills. Group comparison is also valuable because group characteristics used as predictors contribute to our better understanding of the constructs as attainable skills that are influenced (or not influenced) by group characteristics. For this purpose, an ordinary mean analysis or analysis with latent means (which refers to the means at the level of latent variables) can be used.

Second, the present study illuminated the global relation of several selected constructs. A future study could conduct an in-depth analysis of the relationships of these constructs that have only been alluded to here. For instance, an excellent

area for theoretical and empirical research could be (a) a formulation of a more concrete relationship between content and coherence and (b) lexical cohesion, the most uninvestigated area of cohesion.

Third, a future replication may be conducted for students with different language backgrounds and grade levels.

Finally, another contribution of this study is the excellent utility of the writing test instruments: the prompt and the scoring criteria. Researchers can utilize these methods for eliciting and quantifying children's productive language data (written and spoken) to address a variety of topics.

As a final comment, substantively, the theoretically informed empirical investigations carried out in this study make a significant contribution to our understanding of the salient linguistic and semantic resources for text construction, namely cohesion, coherence, content, grammar, and text length. The substance can provide a foundation for conducting subsequent studies such as the ones suggested above. The statistical data yielded by this study and the writing test instruments can make a valuable methodological contribution as resources for undertaking future studies.

ACKNOWLEDGMENTS

This research was conducted partly under the auspices of the UCLA Language Resource Program under the directorship of Dr. Russell N. Campbell. I thank Professors Marianne Celce-Murcia, Lyle F. Bachman, and Russell N. Campbell for their invaluable comments and feedback during the development of this study and on its early drafts. I would like to express my gratitude to colleagues Robert Agajecian and Kathryn Howard for their valuable participation as raters and to Hyesug Lee for her participation in the picture illustrations. I thank the following people for their assistance with the study: Dr. Jamal Abedi, Adeline Shoji, Maureen Hicks, Ruby-Ann Rudnick, Chin Kim, and the teachers in the Korean/English Two-Way Immersion Program and also in the English-only classes. Finally, sincere thanks go to three anonymous reviewers and the *JAL* Editors for their extremely helpful suggestions and rigorous review of the earlier versions of the paper. Any remaining shortcomings are my own responsibility.

APPENDIX A

GUIDELINES FOR TEACHERS FOR DELIVERING INSTRUCTIONS FOR THE ENGLISH WRITING TEST

Please read the basic instructions below and deliver them using your everyday language style and vocabulary familiar to your students. You can add relevant words to these basic instructions flexibly, but keep the basic content given below. Please allow about 10 minutes for these instructions.

Procedures for the Instructions

1. Warm-up:

“Many people would like to know how well you write a story in English...”

2. Example story-writing (Optional):

(Show a two- or three- picture series that has an example story written below it):

“Let’s look at the example story that somebody wrote. (Read the story to the students.) This is only one way of writing . . . You could write differently, e.g., . . .”

3. Actual story-writing task:

“You will NOT write about the example pictures that we just read. You will have NEW pictures to write about. Look at the new pictures.” (Please have the students look at the story line depicted in the 7-picture series for a minute or two. To make sure that the story content is not ambiguous to them, and to activate schemata (background information) for the students, please go over the whole story line with the students, by starting, e.g., “Let’s see what this story is about . . .” Make sure, however, to say the following after you went over the general story: Please do not copy what I just said about these pictures. You MUST write YOUR OWN story; e.g., you may not use the name I used for the boy . . .”

Also, please instruct them about the following:

- * Create your own story/ Write a CREATIVE story, but the story MUST go with the pictures.
- * Your story should be in sequence.
- * Write as long as you can, and as best as you can.
- * Wrong spelling is fine. When you cannot spell a word correctly, you may sound out. Just do your best. Do not pay attention to punctuation.
- * I will give you 30 minutes to finish your writing. It is a good idea to think about what you will write before you begin writing your story on the paper.

APPENDIX B

SCORING CRITERIA

Criteria for Rating English Writing Samples for Early Elementary School Graders (K to 2)

These criteria were developed by Jungok Bae, Kathryn Howard, and Robert Agajeenian.

* These criteria are primarily based on narrative writing: The scale descriptions may vary and are flexible depending on the specific writing tasks and contexts that are changeable.

* Spelling errors, punctuation, and handwriting will not be judged.

* Length: Length will not crucially affect scores in general. Beyond a certain threshold level of length (e.g., once a sample has more than N words), length alone will not be considered a factor to high or low ratings. That is, other qualities will be considered more important than length. However, length is part of fluency, so an essay with words < N will belong to the category of Zero or Limited ability unless the essay is perfectly fine and sophisticated and with no errors. (N will vary depending on the specific task used.)

* The following absolute scale of ability will be used (Adapted from Bachman, 1989). To provide an absolute scale, the two ends of the scale and the scale points in between are defined as follows for Grades K through 2, Grade 2 being the highest grade in this context:

0 _____ 1 _____ 2 _____ 3 _____ 4

0: Zero or very little ability

1: Limited

2: Moderate

3: Extensive

4: Complete: Ideal level of language ability and use for second graders; Characterized by the writing features observed among best second graders' writing samples available and expectations relevant to these student levels.

Ratings with a 0.5 decimal point for each scale point (that is, 0.5, 1.5, 2.5, and 3.5) will be acceptable.

* For further guidelines for scoring, see section "Scoring" in the main body of the paper.

COHERENCE:

Definitions: See section "Construct Definitions" in this paper.

0 (Zero): Too short to judge. No evidence of coherence.
Totally incomprehensible regardless of the length.

1 (Limited): Seriously unconnected/isolated series of ideas.
Serious lack of relationships between ideas.

2 (Moderate): Some connections of separate ideas but no global connections of local ideas.
There may be some major connection missing in between.

3 (Extensive): The whole story organized in general. No serious break. All ideas pretty much connected globally. However, sophistication and elaboration for connections not observed.

- 4 (Complete): Sophisticated and elaborated connection of ideas. Absolutely comprehensible thematically. Ideas absolutely consistent. Clear presence of elements for introductory/opening remarks (e.g., *One day; Once upon a time; Yesterday*), coda, and elaborated connections filled in at every important local point.

CONTENT:

Considerations will be given to relevance, thoroughness, persuasiveness, and creativity.

- 0 (Zero): Too short to judge.
- 1 (Limited): Not thorough at all (Only 15 - 30 % of the content was expressed).
Serious distortion of the picture content.
Large segments of the content missing.
- 2 (Moderate): Somewhat relevant but not thorough.
Some minor irrelevance/inaccuracy.
- 3 (Extensive): The story is complete and thorough in general.
Accurate/relevant in general. In general, FINE, but elaboration and sophistication not observed.
- 4 (Complete): Descriptions of the situations/events just wonderful. Very thorough.
No irrelevance whatsoever. CREATIVE. Persuasive. Convincing.

GRAMMAR:

Grammar refers to morphemes and syntax. Critical errors are defined as errors that seriously impede communication: e.g., a major syntactic chunk missing, incomprehensible word order. Minor errors are defined as errors that do not cause ambiguity in meaning, misunderstanding, and difficulties of communication: e.g., usually errors in morphemes such as third person present suffixes, tense at local level, and plural suffixes.

- 0 (Zero): Too short to judge.
No evidence of grammatical knowledge/use.
No sentences; only single words.
- 1 (Limited): Frequent critical errors. Extensive minor errors.
Few sentences; only phrases.
A sample with length < N words is considered Limited unless the writing contains complex grammatical features.
- 2 (Moderate): Some critical errors. Frequent minor errors.
- 3 (Extensive): Few limitations, no critical errors, occasional minor errors,
with no complex sentences.
- 4 (Complete): Unlimited range. Complex sentences.
A variety of grammatical uses.
Complete control of grammar (Native level). Very few errors.

COHESION:

Definitions and classifications: See section "Construct Definitions" in this paper. Cohesion will be scored based on the number of cohesive markers that are appropriately used.

APPENDIX C SELECTED BEST SAMPLE

Story-making (9/19/21 #271)
20 69

Timmy was walking home after school.
But Timmy saw a girl with a broken umbrella and timmy and you want to come in my umbrella? timmy asked. The girl said are you a stranger? Timmy said no so the girl went in his umbrella and what is your name? timmy asked. My name is linzy said the girl. They walked and talked about things they like. Linzy said this is my house. Then she said since you let me use your umbrella you want to come in my house? asked Linzy. Sure said timmy. Linzy got some papers for her and Timy. You want to draw?

asked Linzy. Good Idea! said timmy. Then Linzy's mom heard that timmy came so she made them some treats and ate until they were full. Linzy's mother said you want some milk? timmy said yes! me too! said Linzy. Timmy said that was so good and I think I'm still hungry said timmy. you want some more? said Linzy's mother. Yes I do said timmy and linzy. Linzy's mother said should I tape the ship on the wall when your finished? said Linzy's mother. Then timmy went home. Linzy said be careful where your going! said Linzy. okay! said Timy. See you next time? said Linzy's mother. I'll come tomorrow

NOTES

¹ Six Korean-American students who were initially identified as English-Proficient were classified into the English-Proficient group in this study to be consistent with the District's classification of students.

² The English oral proficiency levels based on the pre-LAS test are classified as follows: Non, Limited, Functional, and Proficient. Non and Limited are called *Limited English Proficient (LEP)*, and Functional and Proficient *English Proficient (EP)*. Students whose home language is exclusively English are called *English-Only (EO)*; they are also classified as EPs, exempted from the English proficiency identification test.

³ In School A, the school's first and second grade students scored at around the middle percentiles on the CTBS/U subsections against the national norm for the most recent 1994-95 school year (the percentile scores in reading, math, and language were 38, 61, and 48, respectively, for first graders, and 54, 66, and 59 for second graders). All first graders and second graders from this group were identified as English-Proficient, and they were all English-Only students. School B scored in the upper percentiles on the CTBS/U subtests for the most recent two academic years 1993-95 (the percentile scores in reading, math, and language were 65, 89, 70 [1994-95], respectively, and 58, 83, 70 [1993-94] for first graders, and 53, 91, 63 [1994-95] and 85, 95, 83 [1993-94] for second graders). All students from this group, as well, were identified as English-Proficient (EP) on the basis of the pre-LAS (pre-Language Assessment System). However, this group included both English-Only students, whose home language is exclusively English, and non-English-Only students who were proficient bilinguals.

According to the school officials, EP students who are non-English-Only students use English and another language at home; they are proficient bilinguals and treated as EOs in terms of English oral proficiency level. On the other hand, the term English-Only (EO) refers to the students whose home language is exclusively English; thus, EOs are proficient in English but not necessarily in another language (A. Shoji, M. Hicks, & R. Rudnick, personal communication, July, October, 1996).

⁴ For the cases that had a third rating, rater correlation coefficients and reliability coefficients (Table 4) were obtained using the closest two ratings. Six cases, five cases, and four cases for grammar, coherence, and content, respectively, showed discrepancies with more than one scale point difference, and these cases received a third rating.

⁵ Spelling errors, punctuation, and handwriting were not judged in scoring and have been reproduced here as written by the children.

⁶ Due to the high correlation between referential ties and lexical cohesion, some multicollinearity (that is, a high correlation between predictors in using regression) was indicated (an index for tolerance was .187 for both lexical cohesion and reference. To be free from multicollinearity, a minimum tolerance .28 is needed based on the criterion $1 - R^2$, with $r = .85$). However, omitting one of these predictors or collapsing them, which is a typical way to handle a multicollinearity problem in regression, was not applied to this case. Instead, these two variables entered as separate predictors in the regression model because they both concurrently appear naturally in the texts and it is clear that they possess their own attributes of cohesion.

⁷ If X and Y are related, there are several possible explanations: (a) X may cause Y, (b) Y may cause X, and (c) X and Y may be the result of a common cause (Gronlund & Linn, 1990, p. 499). This type of causal explanation is not provided by a correlation coefficient.

REFERENCES

- Arter, J. A., & Spandel, V. (1992). Using portfolios of student work in instruction and assessment. *Educational Measurement: Issues and Practice*, Spring, 36-44.
- Bachman, L. F. (1989). The development and use of criterion-referenced tests of language ability in language program evaluation. In K. Johnson (Ed.), *The second language curriculum*, (pp. 242-58). Cambridge: Cambridge University Press.

- Bae, J. (1997). Cohesion, coherence, content, and grammar in children's written English: The Korean/English Two-Way Immersion Program and single-language (English-only) classes. Unpublished manuscript, University of California, Los Angeles.
- Bae, J. (2000). The construct validation of certain components of English and Korean writing ability in children participating in either a two-way immersion program or monolingual classes: A writing assessment and latent variable approach. Unpublished doctoral dissertation, University of California, Los Angeles.
- Baker, E. L., O'Neil, H. F., Linn, R. L. (1993). Policy and validity prospects for performance-based assessment. *American Psychologist*, 48, 1210-1218.
- Bamberg, B. (1984). Assessing coherence: A reanalysis of essays written for the national assessment of educational progress, 1969-1979. *Research in the Teaching of English*, 18, 305-319.
- Bentler, P. M. (1997). Modeling of intervention effects. *Substance Use & Misuse*, 32 (12 & 13), 1661-1666.
- Berman, L., & Slobin, D. I. (1994). *Relating events in narrative: A crosslinguistic developmental study*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc., Publishers.
- Brown, G., & Yule, G. (1983). *Discourse analysis*. Cambridge: Cambridge University Press.
- Campbell, R., Kim, C., Kim, O., Merrill, C., Rolstad, K., & Bae, J. (1994). *Korean/English Bilingual Two-way Immersion Program: Title VII evaluation report* (Federal Grant No.: T003C20062, 1994-1995), Washington, D.C.: U.S. Department of Education.
- Carrell, P. L. (1982). Cohesion is not coherence. *TESOL Quarterly*, 16, 479-488.
- Collins Cobuild English language dictionary. (1996). New York: HarperCollins.
- Comprehensive Tests of Basic Skills. (1990). Monterey: CTB, Macmillan/McGraw-Hill School Publishing Company (Available by written permission to: Curriculum Frameworks Assessment, CTB Macmillan/McGraw-Hill, 2500 Garden Road, Monterey, CA, 93940).
- Connor, U., & Johns, A.M. (Eds.) (1990). *Coherence in writing: Research and pedagogical perspectives*. Alexandria, VA: Teachers of English to Speakers of Other Languages.
- Cook, G. (1989). *Discourse*. Oxford: Oxford University Press.
- Cox, B. E., Shanahan, T., & Sulzby, E. (1990). Good and poor elementary readers' use of cohesion in writing. *Research Quarterly*, 15, 47-65.
- Cox, B. E., Shanahan, T., & Tinzmann, M. B. (1991). Children's knowledge of organization, cohesion, and voice in written exposition. *Research in the Teaching of English*, 25, 179-212.
- Enkvist, N. E. (1990). Seven problems in the study of coherence and interpretability. In U. Connor & A. M. Johns (Eds.), *Coherence in writing: Research and pedagogical perspectives* (pp. 9-28). Alexandria, VA: Teachers of English to Speakers of Other Languages.
- Fitzgerald J., & Spiegel, D. L. (1986). Textual cohesion and coherence in children's writing. *Research in the Teaching of English*, 20, 263-280.
- Fitzgerald J., & Spiegel, D. L. (1990). Children's writing. *Research in the Teaching of English*, 20, 263-80.
- Freedle, R. O. (Ed.). (1991). *Text and texture: Patterns of cohesion*. Norwood, NJ: Ablex Publishing Co.

- Gronlund, N. E., & Linn, R. L. (1990). *Measurement and evaluation in teaching*. (6th ed.). New York: Macmillan.
- Gundel, J. K., Hedberg, N., & Zacharski, R. (1993). Cognitive status and the form of referring expressions in discourse. *Language*, 69, 274-290.
- Halliday, M. A. K. (1985). *An introduction to functional grammar*. London: Edward Arnold.
- Halliday, M. A. K., & Hasan, R. (1976/1993). *Cohesion in English*. London and NY: Longman.
- Halliday, M. A. K., & Hasan, R. (1989). *Language, context, and text: Aspect of language in a socio-semiotic perspective*. Oxford: Oxford University Press.
- Hatch, E., & Lazaraton, A. (1991). *The research manual: Design and statistics for applied linguistics*. Boston: Heinle & Heinle.
- Herman, J., Aschbacher, P., & Winters, L. (1992). *A practical guide to alternative assessment*. Los Angeles: Association for Supervision and Curriculum Development, UC Regent.
- Kim, C. H. (1996, August). *The Korean/English Two-Way Immersion Program*. paper presented at the Two-Way Bilingual Immersion Summer Conference, Long Beach, CA.
- Koshik, I. (1999). A preliminary investigation into the effect of grammatical cohesive devices: Their absence and their misuse on the comprehension of non-native speaker speech and writing. *Issues in Applied Linguistics*, 10, 3-26.
- Los Angeles Unified School District Information Technology Division, Testing Unit. (1996). *CTBS test results*. Los Angeles: LAUSD Information Technology Division.
- Lindsay, D. B. (1984). *Cohesion in the composition of ESL and English students*. Unpublished Master's thesis. University of California, Los Angeles.
- Linn, R., Baker, E., & Dunbar, S. (1991). Complex performance-based assessment: Expectations and validation criteria. *Educational Researcher*, 20, 15-21.
- McCarthy, M. (1991). *Discourse analysis for language teachers*. Cambridge: Cambridge University Press.
- Mehrens, W. A. (1992). Using performance assessment for accountability purposes. *Educational Measurement: Issues and Practice*, Spring, 3-9.
- Norris, J. A., & Bruning, R. H. (1988). Cohesion in the narratives of good and poor readers. *Journal of Speech and Hearing Disorders*, 53, 416-424.
- Nunnally, J. (1978). *Psychometric theory* (2nd Ed). New York: McGraw-Hill.
- Oller, J. W., & Jonz, J. (1994). *Cloze and coherence*. Cranbury, NJ: Associated University Presses.
- Pedhazur, E. (1982). *Multiple regression in behavioral research: Explanation and prediction* (2nd Ed.). Orlando: Holt, Rinehart and Winston.
- Peterson, C., & Dodsworth, P. (1991). A longitudinal analysis of young children's cohesion and noun specification in narratives. *Journal of Child Language*, 18, 397-415.
- Renkema, J. (1993). *Discourse studies*. Amsterdam: John Benjamins.
- Ricento, T. (1987). Clausal ellipsis in multi-party conversation in English. *Journal of Pragmatics*, 11, 751-775.
- Ripich, D. N., & Griffith, P. L. (1990). Narrative abilities of children with learning disabilities and nondisabled children: Story structure, cohesion, and propositions. *Journal of Learning Disabilities*, 21, 165-73.

- Sacks, H., Schegloff, E. A., & Jefferson, G. (1974). A simplest systematics for the organization of turn-taking in conversation. *Language*, 50, 696-735.
- Schegloff, E. A. (1990). On the organization of sequences as a source of "coherence" in talk-in-interaction. In B. Dorval (Ed.), *Conversational organization and its development*. Norwood, NJ: Ablex.
- Smith, R. (1984). Paragraphing for coherence: Writing as implied dialogue. *College English*, 46, 8-21.
- Spiegel, D. L., & Fitzgerald, J. (1990). Textual cohesion and coherence in children's writing revisited. *Research in the Teaching of English*, 24, 48-66
- Walker, J. F. (1992). Exploring and describing the planning process of a potential Korean/English two-way immersion program with the Los Angeles Unified School District. Unpublished Master's thesis. University of California, Los Angeles.

Jungok Bae recently received her doctorate in applied linguistics at UCLA. Her research interests include language assessment and language acquisition and related applied linguistics topics and research methods.

Two Interviews on Language Testing: An Introduction

Nathan Carr

University of California, Los Angeles

Viphavee Vongpumivitch

University of California, Los Angeles

The following section contains interviews with noted language testing experts Charles Alderson and Dorry Kenyon, whom we had the chance to interview during the Fourth Annual Southern California Association for Language Assessment Research (SCALAR) Conference held in Los Angeles on May 11-12, 2001. The theme of the conference was "Foreign Language Assessment at School and College Levels," an area in which both Alderson and Kenyon have much experience and insight. They provide complementary perspectives on issues in language testing because of their varied backgrounds, research interests, and the different test development and research projects with which they have been involved.

Alderson, an applied linguist by training, is a professor of applied linguistics at Lancaster University. He has done a great deal of work on both theoretical and practical aspects of language testing and applied linguistics research, primarily in Europe and the British Commonwealth. His work includes research on language test design methodology, test validation, and the assessment of reading. He is Scientific Coordinator of a Web-based diagnostic language testing system (DIALANG), sponsored by the European Union and conducted in 14 different languages.

Kenyon, on the other hand, was trained as a quantitative methodologist who has always strived to apply these methodologies to testing language. Rather than pursuing a career in academia, he works as a researcher and test developer at a nonprofit research center, The Center for Applied Linguistics, in Washington, DC. He has done a great deal of work developing tape-mediated Simulated Oral Proficiency Interview (SOPI) and computer-mediated Computer Based Oral Proficiency Interview (COPI) versions of the American Council on the Teaching of Foreign Languages Oral Proficiency Interview (ACTFL OPI), the primary tool for assessing foreign language speaking proficiency in the United States. He is also currently working on the development of the new foreign language section of the National Assessment of Educational Progress (NAEP), sometimes referred to as the "nation's report card," as well as on developing web-based proficiency tests of less-commonly taught languages.

Because of the range of perspectives they provide, these two interviews together yield valuable insights into a number of concerns central to language testing and research today. We are grateful to Dr. Alderson and Dr. Kenyon for agreeing to talk with us about their ideas and experience.

An Interview with J. Charles Alderson

Viphavee Vongpumivitch
University of California, Los Angeles

Nathan Carr
University of California, Los Angeles

PROFILE

For more than 20 years, J. Charles Alderson has been an internationally respected scholar in language testing. He has published research in a wide variety of areas, including reading assessment, test development, test validation, test impact (*washback*), computer-based testing, English for specific purposes testing, the effect of background knowledge on student performance, and the relationship between language testing and second language acquisition theory. His most recent book, *Assessing Reading* (Alderson, 2000), explores the nature of reading ability as well as issues involved in constructing and evaluating reading tests. Dr. Alderson is also a co-author of *Language Test Construction & Evaluation* (Alderson, Clapham & Wall, 1995), and he is also co-editor of the journal *Language Testing* and the Cambridge Language Assessment Series. Dr. Alderson is currently Professor of linguistics and English language education in the Department of Linguistics and Modern English Language at Lancaster University, United Kingdom. He is also the Scientific Coordinator of the DIALANG project, a web-based diagnostic test of 14 European languages, and he is an advisor to the British Council on the Hungarian English Examination Reform Project.

INTRODUCTION

The interview is divided into six sections. In the first section, Dr. Alderson briefly explains how he became involved in various areas of research. In the second section, *Washback Research*, Dr. Alderson discusses his work on the impact of tests on classroom teaching and learning, as well as teaching materials development and education policy. Dr. Alderson and Dr. Dianne Wall were among the first researchers to initiate systematic investigation of positive and negative impacts that tests may have on education. From their field research in Sri Lanka, Alderson and Wall (1993) proposed the groundbreaking *Washback Hypotheses* which describe the nature of washback and identify the types of influence that a test can have on both teachers and students. In this interview, Dr. Alderson reflects on the effort to create a theory of Washback in Alderson and Wall (1993), the need for

empirical studies, and the role of Washback in his current research in Hungary on English education reform.

Shifting to another area of research, we then ask Dr. Alderson about the challenges, advantages, and disadvantages of computer-based and web-based language testing in the third section of this interview, *Computer-based and Web-based Testing*. At present, there is a growing interest in using computers and the Internet as media for delivering tests (see, for example, *System*, Vol. 28, No. 4, a special issue on technology and language testing). Based on his experience as the Scientific Coordinator of the DIALANG project, Dr. Alderson discusses the issues involved in developing web-based tests. He urges language testers developing computer- and web-based tests to avoid the use of traditional test item types such as multiple-choice and cloze, and suggests they move toward creating more useful kinds of test items that will make the best use of the new testing medium. He also discusses other challenges to web-based testing, such as overcoming the lack of security and finding ways to effectively score open-ended items, and the challenges faced by emerging research on ways to incorporate corpus analysis into language test development.

In the fourth section, *Test Validation and the Quantitative/Qualitative Paradigms*, Dr. Alderson shares his views on the characteristics of effective language test validation research and addresses issues surrounding the commonly discussed dichotomy between quantitative and qualitative research. Dr. Alderson challenges the either/or nature of this purported distinction, emphasizes the need to integrate both paradigms, and calls for more collaborative study by teams of researchers with different specializations.

In the fifth section of the interview, Dr. Alderson discusses second language reading research. He argues that even though much research has been conducted over the past 30 years, remarkably little is known about this area. Dr. Alderson holds that it is still difficult to make definite conclusions about any particular area of reading research except the role of background knowledge in reading comprehension.

Finally, in the last section we ask Dr. Alderson for suggestions for beginning researchers about conducting language testing research and about good introductory level books for those interested in learning more about language testing.

THE INTERVIEW

Carr: *How do you view the relationship between the various areas of research that you've conducted over your career?*

Alderson: I consider myself an applied linguist first and foremost. Obviously, I'm interested in language testing, but I'm interested in all things to do with language use and language learning, particularly in the second language context. To do testing you have to understand the construct you're trying to test. Pretty much

everything that I have done relates to teaching problems, one way or the other, and how to evaluate them.

Vongpumivitch: *You have said that your early work stemmed from problems you faced as an ESL teacher. Now that you are no longer an ESL teacher, what is it that sparks your interest in research projects?*

Alderson: Often what happens is somebody with a problem to solve comes along and says, "Can you help?" So back in '86, the British Council came along and asked if I could help them revise the ELTS (English Language Testing Service) test, which was the old British Council test. They wanted to bring it up to date with applied linguistic theory, improve it, and make it shorter. I said OK and became director of the project, and we created the IELTS test (International English Language Testing System), which you've probably heard of. It was reasonably successful. It was an interesting test, and we wrote some interesting research articles as a result of that. So that was one example. In Hungary, where I've been living for the last two years until last summer, one of my jobs was to advise on exit-exam reform. They wanted to help people develop decent tests and to train them in testing-related matters. So again, people asked me if I could help.

Washback Research

Vongpumivitch: *Your article with Dianne Wall (Alderson & Wall, 1993) argued that tests have an impact on classroom teaching and learning and explored the nature of such impacts, called washback, as well as ways to measure them. How did you become interested in studying washback?*

Alderson: The notion of washback came out of Sri Lanka. I was involved in Sri Lanka because we at Lancaster, in the Institute for English Language Education where I was director at the time, had teachers and teacher trainers coming to Lancaster to improve their understanding of language education. We were involved in advising partly on test reform and teacher-training reform, and through that connection they asked us if we would get involved and advise them on creating a new school-leaving exam. What we were particularly interested in doing was proving that washback works. So we got some money from the British government and designed a study that lasted about three years to prove that improving the tests as well as improving the textbooks and the teacher training would have the impact required. That was really the agenda. Unfortunately we were wrong. We showed that there was washback on content of teaching, but not on methodology of teaching, and that's where the article came from—the surprise of the lack of success of that project.

Carr: *Methodology would probably be the hardest thing to get a language teacher to change.*

Alderson: That's right, because it involves changing the way they think, basically.

Vongpumivitch: *Beliefs about teaching exist at the individual level, which may be the hardest place to bring about change.*

Alderson: Yeah, it probably is. I mean the stuff we've done since looking at TOEFL, for example, does show some impact on methodology, but it's not predictable because what we see is that teachers are very much influenced by their own teaching styles which relate to personality as well as their experience from the past. Very often people teach the way they were taught.

Vongpumivitch: *Most washback studies seem to be reports of case studies. What can we do with this information? At the end of the day, how do you make sense of the whole field of test impact research?*

Alderson: Read the Alderson and Wall paper (1993) where we talked about the Washback Hypotheses. Those hypotheses are the beginning of theorizing about impact, going beyond the primitive assumption that tests have negative impact to a more general statement that tests will have general influence and not necessarily a negative influence. So at that level, those Washback Hypotheses are a theoretical framework, and the case studies are ways of exploring that theoretical framework.

What we suggested in that article was that there are two theories from outside applied linguistics that we need to understand: innovation theory and motivation theory. More work has been done looking at innovation theory than has been done at the individual level of motivation. We were thinking of students' motivation when we wrote that article. But since then, I've come more and more to realize that we also need to understand teachers' thinking and cognition. In other words, what are teachers' belief frameworks? Why do they do what they do? What drives what they do? And so washback studies, I think, can potentially contribute to studies of teachers' thinking and theories of teacher cognition, as well as draw from research on teacher cognition in order to explore reasons why teachers do what they do. There are bits and pieces in a couple of papers I've written but I've not fully developed anything yet. A lot of people who have been working in teacher training and teacher education are doing that kind of research already. But one of the interesting things about a lot of what happened, particularly in Britain, is that it's not very empirical. It's argumentative; it's assertive; it's theoretically oriented. But there is not very much research that actually goes into the classroom, interviews the teachers, and asks why they did what they did. It tends to be the washback research where we do that sort of thing.

Vongpumivitch: *So there's definitely a need for empirical study.*

Alderson: Oh, yeah, absolutely.

Carr: *Do you think that it will be useful for an empirical study to come up with a general definition or sets of criteria for assessing the effects of washback?*

Alderson: What I've done in relatively recent work, before the Hungarian stuff, was to try to predict washback; that might be what you mean by criteria. So for example, for the University of Cambridge Local Examination Syndicate, several of my students and I designed a set of instruments to investigate the impact of IELTS. And the way we did that was by predicting what a positive influence of IELTS on classrooms would look like, in terms of the usual content—text types, task types, skills—but we also tried to predict what sort of methodology teachers would be using in the classroom. We can turn our results into a framework for doing washback research on IELTS and developing instruments for conducting washback studies, such as classroom observation schedules.

Carr: *So it would be more situation dependent. It wouldn't necessarily be one that can be generalized to all situations, just like there cannot be one test that can be used in every situation.*

Alderson: There are probably generalizable elements you could arrive at once you've done the particular. I think we now understand more from the case studies that have been published than we did before. Those case studies were embedded in individual contexts, so we can generalize from the mass of case studies, and I think methodologically we're beginning to understand what sorts of things we might want to predict regardless of the tests. In other words, we want to look at test constructs, test contents, test methods and predict from those.

Vongpumivitch: *Have you seen any example of a successful washback driven test reform?*

Alderson: The work we are doing in Hungary, I hope, will be one example. It hasn't happened yet.

Vongpumivitch: *You have briefly mentioned that you are currently involved in an education reform project in Hungary. Could you tell us more about it? Isn't it a teacher-training project?*

Alderson: There's a lot of teacher training going on in order to engineer positive washback. What we've developed is a test design which now has been piloted in three rounds. The test has not yet been introduced into the system and won't be

introduced for another four years. We have been working on a project in collaboration with the Hungarian authorities to develop tests which are communicative. The tests relate to the sort of teaching teachers say they want to do and are different from what they say they don't want to do. The difficulty is that the tests must also be for credit, which means that some students will fail, and the tests are therefore high-stakes tests and can have negative impact if teachers don't understand what is required by the test. So parallel with the technical things we are doing like standardizing, standard-setting, and piloting, colleagues have been developing an in-service teacher training course to help teachers understand what the test is about, why it's the way it is. They are trying to get teachers to think about how they would teach in class when they are given a test like this. For example, if you're going to teach reading in an exam preparation class, then what's the best way to teach reading in general, or listening in general? So it's relating best test preparation practice to best classroom practice.

Vongpumivitch: *It seems that if the Hungarian project is successful, it can be a model that other countries can use.*

Alderson: That's what we hope, certainly. We now have three publications; the third volume has just come out, giving details of the teacher-training course. Two books have been published already called *English Language Education in Hungary: Baseline Study* and *English Language Education in Hungary, Part II: Examining Learners' Achievement in English* (Alderson, Nagy, & Oveges, 2000). Part III is about the in-service course that we will develop, and is subtitled *Training Teachers for New Examinations*. It is edited by Együd Györgyi, Gál Ildikó and Philip Glover and is available from Edit Nagy in the British Council, Budapest.

When doing a washback study, you have to be careful. There's always a danger that teachers will fool you. They'll do things that are unprofessional as a shortcut. Teachers vary enormously, as you know. Some think seriously about what's the best way to prepare their students in general as well as for tests, and some don't. Particularly in countries like Hungary or Sri Lanka where teachers are not well paid, they have to go and do other jobs. They take shortcuts, just as they do here in America when teaching TOEFL. So one of the tricks, I think, in engineering washback is to think how teachers can possibly subvert the test and then try to find ways around that subversion. A lot of people don't like such an approach. When people talk about teaching, they tend to talk about teachers as if they were essentially professional people who want the best for their students, but not all teachers are like that. So you have to think of those teachers and how they might teach—you have to consider the worst case as well as the best case.

Computer-based and Web-based Testing

Carr: *What do you see as some of the advantages, disadvantages, limitations, and doors that are opened by computer-based testing and web-based testing?*

Alderson: The biggest danger of computer-based testing in general is that the methods used in the tests tend to be conservative. After all these years, we still have multiple-choice, and we still have gap-filling or cloze tests. So it's a real worry that the test methods will be conservative compared to what we do in paper-and-pencil tests or face-to-face tests. But with the web-based tests and with the increasing power of IT in general, I think people are starting to experiment with method; for example, the TOEFL listening test is enhanced by having visuals. We don't actually know whether it improves the test, but my hunch is that it does. Similarly when on-stream video becomes widely available that could also have an advantage.

The biggest disadvantage of web-based testing for many situations, though, is not the method, but the lack of security. Having secure web-based tests is problematic. So the TOEFL, for example, will never become web-based, I don't think, at least not in my lifetime. Where the most interesting developments in web-based testing or computer-based testing are going to happen is in low-stakes environments rather than high-stakes environments such as placement testing, diagnostic testing, classroom testing. The problem is most money doesn't go into producing classroom tests, or into producing an achievement test or if it does, it tends to become a high-stakes test. You need to have resources to be able to develop interesting test methods.

Carr: *There are placement tests here [at UCLA] that are moderately high stakes, and we are trying to take them web-based.*

Alderson: It will be interesting to see what you're going to do. Internet-based? Well, security will be a problem. Personation will be a real problem. How do you guarantee that the person who's taking the test is the person who is applying to come to your classes?

Carr: *For now we're planning on lab-based administration.*

Alderson: Right. There you go. Which means that you're eliminating the advantages of web-based testing. It no longer individualizes but it's group-based.

Vongpumivitch: *Are you saying that very high-stakes tests should be paper-and-pencil as opposed to web-based?*

Alderson: I just think that the advantages become a bit more debatable. Obviously rapid feedback is very useful, with the web-based test. Having a record of scores and so on in terms of research purposes as well as administrative purposes is very handy. It cuts out the stage of having machine readable answer sheets. So there are advantages but there are not very many construct advantages.

Vongpumivitch: *Beyond improving test administration processes, how can a computer-based or web-based test provide better interpretation of the test scores? How can we have a better measure of language proficiency using computers or the Internet?*

Alderson: I think the incorporation of multimedia is an interesting possibility because you can do all sorts of simulations. Ultimately your test method is almost certainly going to be selective-response type method, which has problems, but you can see how scenarios could be developed for more integrated skills testing, in the receptive skills at least.

Carr: *So you don't see much of a role in the immediate future for automated scoring of open-ended, limited production tasks.*

Alderson: Well as you know, some work on that has been going on at ETS, but I think that's several years down the pike before artificial intelligence progresses far enough to deal with English as a second language. A lot of stuff that they are doing in developing algorithms is English as a first language, and they are having some degree of success.

Carr: *They are doing that for essay tests. But what about specifically for the short-answer type—one word, one sentence?*

Alderson: There are people at ETS, Jill Burnstein is one of them, who believe that they can develop artificial intelligence systems to do that, but I haven't seen that being done successfully.

Vongpumivitch: *But then you only need artificial intelligence in the case of ETS tests, because there are so many people taking their tests. But if your test taker pool is not that big, then the need for artificial intelligence may not be as strong.*

Alderson: Well ETS has the resources to develop artificial intelligence systems, if they would then make them available to the education community that would be great. The size of your test taking population does not matter so much, I don't think, provided you are dealing with the same language and essentially the same construct and test methods.

The other great hope, of course, is corpus-based testing, where you use your corpora to provide criteria for deciding whether something is acceptable or unacceptable. The sorts of corpora I'm thinking of are the sorts that have been developed particularly in Britain—the Bank of English and the British National Corpus, which are most interesting in their written forms, not in their spoken forms. Concordancing is a powerful tool which can be used on corpora. Imagine you have a structural pattern, or frame, in which you want to create an item. One could search in the corpus for that pattern or frame and concordance elements that occur within that frame. The problem is, no matter how big your corpus is, it may not contain the frame that you want to use in your items. So presumably what you then have to do is to take the language from the corpus in order to produce the frame, and then go back to the corpus to get the range of possible responses. With a parsed and tagged corpus, you get information about parts of speech, about the syntax, and you could search the corpus for examples of a particular structural framework.

Carr: *Could you elaborate on the notion of structural frames a little bit?*

Alderson: What's been developed in Lancaster alongside the British National Corpus is the *claws-tagging* which also gives you some information about grammatical function: not just the part of speech, but also the subject, object, verb, that sort of function, the fairly basic aspect of clause structure. That's why it's called *c-l-a-w-s*. You can search for particular syntactic features associated with particular parts of speech. For example you can ask for examples of a pronoun in a syntactic frame, and so on. So you could use that information to construct tests, if that's the sort of test you want to construct. Obviously it will be a test of fairly low level linguistic information. People are working on anaphora marking, and there are two people in Lancaster who have been working on semantic tagging, developing semantic frameworks for identifying meanings.

Vongpumivitch: *But those sound like grammar tests to me.*

Alderson: Of course that may be another reason why you might not be interested, that's right. But there are a lot of things you could do. Any decent corpus will have texts, classified by type, genre, so you could go to your corpus and say, "Give me a text of the following type, on the following topic, with the following structure."

Vongpumivitch: *And you'll get your reading passage right there.*

Alderson: That's right. You then have to go ahead and start constructing items that tests whatever skills you want to test. That's something the computer can't do for you.

Vongpumivitch: *But the answers to those questions, if they happen to be open-ended questions, cannot be checked by the corpus.*

Alderson: Correct, unless we've got a semantic-parsed corpus, and some people have been developing parsing systems that will, for example, suggest synonyms, paraphrases, and so on, of a particular word.

Carr: *At this point you're getting into artificial intelligence.*

Alderson: Yeah, it's getting close.

Carr: *What do you think should be some of the priorities in the development of and research on computer-based, web-based testing?*

Alderson: Whenever I talk about DIALANG, for example, or computer-based testing more generally, people who aren't very keen on information technology ask the question, "What is the added value? What do you get from doing this that you couldn't get using paper-and-pencil?" And obviously some of the answers are practical ones. I think if you can show that you're enhancing the construct, that you're enhancing in some sense the validity, reliability, or the usefulness of the test, then I think that's where research should go, preferably on the validity side. The reliability side is fairly obvious, I think, and usefulness is not always clear. I think the other area to look at is negative impact. What will be the impact of computer-based tests compared to the paper-and-pencil-based test? And nobody has done that kind of research for TOEFL, for example.

Carr: *Do you mean that nobody has looked at how computer use is going to disadvantage some students?*

Alderson: Well, people have looked at the disadvantages, of course. Computer literacy was examined in the initial study for TOEFL CBT. Studies showed some disadvantage in some parts of the population. What they didn't really show was how people with and without computer literacy perform on the new TOEFL, rather than on TOEFL-like tasks. What ETS did is rather different. What ETS did was develop technology familiarity questionnaires that are administered during the regular TOEFL administration. It's not the same thing as looking at the impact of computer literacy on TOEFL CBT. But interestingly, ETS has never done any research into TOEFL washback. There are no TOEFL washback studies, apart from the little thing that I did with Liz Hamp-Lyons (Alderson & Hamp-Lyons, 1996). There should be TOEFL washback studies. There are so many complaints out there about TOEFL. It's unprofessional of people not to have studied its impact.

Vongpumivitch: *But it will be a study of impact on what?*

Alderson: On the students, on how they prepare. On the teachers, on how they teach. The study we did (Alderson & Hamp-Lyons, 1996) was in the United States. We looked at two teachers, teaching their regular classrooms and their TOEFL classes. We looked at impact on the teachers and how each teacher changed what he or she did in the classroom.

Vongpumivitch: *But the majority of people preparing for TOEFL go to test prep schools.*

Alderson: OK, then let's go to the test prep schools and see what they do. They guarantee to increase your scores by 60 points. How do they do that? What is it they're teaching? Why are they teaching that? Why don't they teach something else? It will also be useful. Nobody has looked into it. Why not?

Test Validation and the Quantitative/Qualitative Paradigms

Vongpumivitch: *In your opinion, what's the primary goal of a test validation study? What should be the characteristics of good language test validation research?*

Alderson: What you want a validation study to do is to show that the test you are studying, that the inferences that you've drawn from the test scores, and the uses for the test scores are justified. That's tricky, because what you would like to be able to do is to show that you can make better inferences on one sort of test as opposed to another sort of test. In other words, very often test validation is most useful when it is comparative. Most test validation isn't. Most test validation is also problematic because it tends to be with truncated samples. It tends to look only at those students who had succeeded in some way. Look at predictive validation studies, for example, how do you do a predictive validation study with TOEFL without letting in the students with low scores? The only people you look at are those who got in. So ideally a validation study will be used before the test becomes operational; it will look at the full range of possible consequences of the scores before we use the scores from that test. Most validation studies are much more limited because of the truncated samples and because of the nature of the criteria against which they are comparing the test scores. Indeed they are often limited because of the rather quantitative nature of validation studies. Ideally a validation study is both quantitative and qualitative. For example, I have a student, Jay Banerjee, who's looking at IELTS and its predictive validity. What she's been doing is looking in great detail at how admission officers actually use IELTS scores, how they make decisions by taking into account the IELTS scores, as well as all the other information. Then she's interviewing, again in great depth, the individual students whose scores have been used, looking at the problems they have in their study setting and trying to understand the complexity of their language use prob-

lems. Ideally, a test validation would look in considerable depth at those issues, for students who have got low scores as well as those who have got high scores.

Carr: *Some people criticize language testing research as overemphasizing quantitative or psychometric methodology, at the expense of qualitative methods. But on the other hand, others criticize qualitative methodologists for focusing too much on case studies which are not very generalizable. As students we were told in our research methods classes that the ideal approach involves a complementary use of both paradigms in a study. How realistic a goal do you see this in language testing research?*

Alderson: As always it depends upon the purpose of the validation research or whatever it is you are doing, and it depends upon the resources that you have. To begin with, I don't necessarily see that there is an essential difference between quantitative and qualitative research. It is clear that qualitative researchers have to quantify, because they use words like *some*, or *many*, or *exception*, and so on; that's quantification. And, similarly, quantitative researchers are concerned about the quality of their instruments; that's what validation is. So this is a false dichotomy. I think generally it's accepted now in the social sciences that the most sensible approach is a fusion of both. If we stay with the words for the moment to understand the differences, qualitative research is very resource hungry, not only in terms of gathering the data—you have to interview people, observe them, or whatever you're doing—but also in terms of analysis. Analysis is extremely time consuming with the sort of data you have, even using software. There are good packages out there, but they don't get the grasp of what you have to do in coding. That's why qualitative research is either not very well done, or is only done superficially, or as an afterthought to quantitative analysis, because quantitative results are more amenable to analysis. Of course, quantitative data is relatively easy to gather. People have to take a test anyway. You just give them questionnaires as well, and you can get the data quickly. It's hard to get data for qualitative research. Ideally you need both. And a lot of the stuff that I've done has been in fairly small numbers and more toward the qualitative end than the quantitative end, typically because Britain is a smaller place; we have smaller populations to deal with.

Carr: *What do you see is the best way to go about integrating the two paradigms?*

Alderson: I think the value of qualitative research is that it helps you understand the problem, or identify the dimensions of a problem. So you could do the qualitative stuff first as a way of piloting to get inside the complexity of the situation, and then possibly follow up with analysis to identify key variables which might be worth exploring in greater extent or in more depth, and follow that up with quantitative studies. But I think both qualitative and quantitative researchers are worried about generalizability. What is generalizability? Very often, to generalize means

to ignore important variables. We know how important context is, and we know how situation-dependent most of our work is. So I think you have to state the limit, as you understand it, to the generalization you're making.

Carr: *Are there any good examples you can think of in language testing that have succeeded in using both approaches?*

Alderson: I guess the classic one is the paper by Anderson et al. (Anderson, Bachman, Perkins & Cohen, 1991) —the triangulation study. That's a nice paper. It's a nice example of how you can go about examining the same problem from different perspectives. I guess it's true to say that the TOEFL-Cambridge comparability study (Bachman, Davidson, Ryan & Choi, 1995) is not a bad example of an attempt to use content analysis in a quantified way in order to shed light on the quality of the instruments involved. There were still more data they could have gathered, introspective data for example, but I think that it was a good study that was well resourced and took a long time, longer than a graduate student can possibly have done alone.

Vongpumivitch: And it has to be a team effort.

Alderson: Absolutely. See the big problem with a lot of testing research is that it's done by graduate students. You need a team of researchers, and that isn't recognized for the award of a PhD. So we should be having more funded research, teaming people with different skills, taking place over time. Most research that we do is one shot research rather than developmental, and we should be doing more developmental research. But that takes time. We should learn from the sciences. What we should learn is to replicate. Give people PhDs if they replicate adequately because a PhD in our field is an apprenticeship to do research; it qualifies you to be a researcher. So limit what you demand of somebody at the PhD level. Don't expect original, creative research. Do that later. Once you've got your PhD, team people up. That's the way it should be done.

Second Language Reading Research

Vongpumivitch: *You've been in the field of reading for about 30 years now. Reading is a field that is heavily researched. What is it that we know enough about now, and what is it that we still don't know?*

Carr: *What are we sure about?*

Alderson: We're pretty sure it's complex. We're pretty sure people can do it but we don't know how. I often wish I never got into it because I'm not sure what I've learned as a result of 30 years of research.

Vongpumivitch: *Seriously? Even after writing a book about it?*

Alderson: Well, think about the stuff I've done on skills, and again I started on that because of a particular real-world problem that somebody came up with. I was in India in 1986 and somebody was saying, "What we're doing is using Benjamin Bloom's taxonomy to test our students' ability in English," and I said, "You can't do that. Native speakers are different, and so why are you doing that with nonnative speakers?" So the research I set out to do, looking at levels of skills, was to prove them wrong in India. And I proved them wrong, but I'm not sure if I've proved anything beyond that. Everything I've done since then in looking at skills has convinced me of the complexity of the issue, but hasn't reached the solution. As I said in the book (Alderson, 2000a) basically it seems to me that what happens is individuals use strategies and skills in individual and idiosyncratic ways, depending upon purpose and knowledge, etc., and it's very hard to generalize from what individuals do to some deeper understanding of what is involved. Of course we have schema theory but it has got us nowhere. We know background knowledge has an effect and that's hardly a surprise.

Carr: *How compensatory are these varying ways?*

Alderson: I hate to say it, because I was brought up as an applied linguist in the 1970s when we believed that skills and strategies are important, but I think in ESL or French as a foreign language, the language is what you need before you can have adequate transfer. So the threshold hypothesis is still very important. The reason why I say "I hate to say this," is that I fear for the washback of statements like "You need to know the language first," because what people start doing when they teach grammar or vocabulary may not be the best way to learn a language. But I'm certain you need a good linguistic foundation because then you can start all these other things. Now that doesn't help very much. Frankly if you had sat me down 30 years ago I probably would have said the same thing. So what have we learned after 30 years? I don't know.

Vongpumivitch: *Is there any hope for reading researchers at this point? Are there any areas that need to be investigated?*

Alderson: Every area, you name it, can be investigated. Absolutely. I've got a student at this moment investigating reading aloud. I thought that had died out 30 years ago, but she still finds some interesting problems about it. It's all up for grabs, and I'm not sure I want to supervise it anymore. It's very frustrating. *Assessing Reading* is a long book. It caused me a lot of sweat to write that book. I had to read an enormous amount, and I don't think there's a clear message coming out of that book.

Vongpumivitch: *At the end of the book you still had to conclude that this is not a conclusion.*

Alderson: That's right.

Carr: *Is it fair to say that you think the types of research questions that have been addressed over the years haven't really evolved so much, that they have come full-circle? Are we still looking at the same issues as 30 or 40 years ago?*

Alderson: I suspect we're not, but we should be. Obviously we have a better understanding of terms like *strategies*, and even *skills*. We have a better idea of how we can do research, through introspective research, for example. But one of the problems with the field is, what we don't do is, build on each other's research. So what we're not doing is saying, "OK, let's develop a program of research that explores these different angles, and accumulate knowledge." What happens is that people are doing their own things. This goes back to what I was saying about graduate students. Doing research on your own is great, but it's fragmented.

Closing Thoughts

Vongpumivitch: *There's a perception on the part of many people outside the language testing community, that the field of language assessment is strictly quantitative and generally incomprehensible and inaccessible to non-specialists. What is your response to the people who say they want to come into testing, but feel that it is too inaccessible to them?*

Alderson: Applied linguists should know that they can learn enough of the quantification in order to throw light on the problem that they are trying to address. You don't need to do all these fancy things in order to address particular problems. Statistics are there to be used as a tool to help you understand something, and not vice versa.

You could do discourse analysis; look at the studies into OPIs and SOPIs and all the other language testing studies that utilized discourse analysis. Some of the work has been done; Steve Ross and others, for example, have looked at interlocutors' accommodations (Ross & Berwick, 1992). Ross' study is a decent discourse analysis that throws light on how we might improve the training of interlocutors.

Vongpumivitch: *How can a graduate student or a novice researcher start a language testing study?*

Alderson: Start with a problem, and see how testing can help. This was why I got involved in testing. I was told to develop a placement test in my very first teaching

job. What was the problem? To identify weaker students from stronger students and put them into homogeneous groups. So I had to get involved in testing in order to address that problem. That's where interesting testing work comes from—having a real-world problem, then trying to do research.

Carr: *What are some introductory level books that you would recommend for someone who is interested in language testing?*

Alderson: Tim McNamara has a nice new book in the OUP series (McNamara, 2001) It's a nice introduction. Arthur Hughes' book on testing for teachers (Hughes, 1988) is pretty superficial, but it's a good start. Brian Heaton's second edition, *Writing English Language Tests* (Heaton 1988), is a nice easy introduction. Going beyond those, you can get Cyril Weir's book on standardizing and developing tests (Weir, 1993); that's a good book. The book I did with two colleagues (Alderson, Clapham & Wall, 1995) is a bit more technical but people find it readable. If people really got hooked by then, then obviously they should buy all the books in the series by Lyle Bachman and myself, which is learning about constructs. The best way of finding out what vocabulary is in Applied Linguistics is to read John Read's book about vocabulary (Reid, 2000), because he tells you about the construct as well as testing. That's what those books are intended to do, to show the centrality of testing to applied linguistics. It's not peripheral; it's central.

SUMMARY AND CONCLUSION

In this interview, Dr. Alderson shares his opinions and gives suggestions on four key areas of research in language testing: test impact (washback), computer-based/web-based language testing, test validation research, and testing reading comprehension. He emphasizes the need for empirical studies of test impact, especially the investigation of teachers' motivation and the prediction of washback. In terms of computer-based and web-based language testing, Dr. Alderson argues that it is crucial to show that technology enhances the ways language ability can be measured and helps create tests that have better quality and yield better score interpretation. He also discusses the disadvantages of computer-based and web-based language tests, such as the lack of test security and the difficulty of grading open-ended items using complicated artificial intelligence.

Dr. Alderson advocates more cooperations among researchers from different disciplines, arguing that more team research will enrich the quality of language testing studies. He believes that the best test validation research is that which incorporates both quantitative and qualitative research methods, investigating all related issues in great depth. Cooperation is also needed in the area of reading assessment; Dr. Alderson points out that there are still many unanswered questions in all aspects of reading assessment.

Perhaps the most important point in this interview is Dr. Alderson's emphasis that language testing is a central field of study in applied linguistics. Dr. Alderson explains that the perception of language testing as a strictly quantitative area of study is a misunderstanding. Language tests are measurements of language abilities, and language testers need to first have a clear understanding of the nature of language abilities before they can measure them well. Dr. Alderson argues against the notion of a dichotomy between qualitative and quantitative methods since few studies are strictly one or the other. Ultimately the research questions motivated by real-world issues are the language testers' guides, and various kinds of quantitative and qualitative research methods are merely tools for language testers to use to conduct an applied linguistic study.

NOTES

¹ The interview took place at UCLA while Dr. Alderson was at Los Angeles as one of the invited keynote speakers for the Fourth Annual Conference of the Southern California Association for Language Assessment Research (SCALAR 4) at the California State University, Los Angeles, CA, May 11-12, 2001.

² Currently there are four books in the Cambridge Language Assessment Series: J.C. Alderson (2000a) *Assessing Reading*, G. Buck (2001) *Assessing Listening*, D. Douglas (1999) *Assessing Languages for Specific Purposes*, and J. Read (2000) *Assessing Vocabulary*.

³ DIALANG is a project funded by the European Union for the development of diagnostic language tests in 14 European languages. Tests will be made available on the Internet free of charge. The DIALANG project will offer separate tests for reading, writing, listening, vocabulary, and grammatical structures, covering all levels from beginning to advance. For more information on the DIALANG project, please go to: <http://www.dialang.org>.

⁴ Dr. Alderson taught EFL and applied linguistics in Germany, Algeria, Scotland, and Mexico. He became interested in language testing and applied linguistics when he was an EFL teacher in Germany, which led him to his postgraduate study at the University of Edinburgh. His early works, such as works on reading, cloze tests, and metalinguistic knowledge stemmed from teaching-related problems he encountered as an EFL teacher.

⁵ The International English Language Testing System (IELTS) is an academic English as a Second/Foreign Language test required by Australian, British, Canadian, and New Zealand post-secondary institutions. All non-native English-speaking students wishing to enroll in such institutions have to take this test, which consists of listening, speaking, reading, and writing, from the University of Cambridge Local Examinations Syndicate, UK. Nowadays, some American post-secondary institutions accept the IELTS scores in place of the Test of English as a Foreign Language (TOEFL) scores.

⁶ *English Language Education in Hungary, Part II: Examining Learners' Achievement in English* is a collection of progress reports on the Hungarian English Examination Reform Project co-edited by Alderson.

⁷ In language testing, a construct is a definition of the ability that is to be measured by the testing instruments. Such a definition of language ability needs to be appropriate to the particular testing situation, test purposes, test taker population, and types of actual language use in the real world (Bachman & Palmer, 1996, p. 66).

⁸ According to Biber, Conrad, and Reppen (1998) and Kennedy (1998), to conduct a corpus linguistic analysis, a corpus user can use software to display all occurrences of a search item, such as a keyword or a syntactic morpheme, in a corpus. A concordance is a formatted display of an exhaustive list of all of the occurrences.

⁹ This study on reading comprehension tests investigated the relationship among test taking strategies, content of test items, and the students' test performance by using think-aloud protocols, content analysis of each test item, and statistical analysis of the think-aloud protocol data and the test performance data.

¹⁰ In this study, Bachman et al. investigated the comparability of the TOEFL test and the First Certificate in English (FCE) test, which was created by the University of Cambridge Local Examination Syndicate, by gathering data from ESL students in eight countries around the world and conducting statistical analysis of the test performance data along with expert judges' ratings of the two tests' content.

¹¹ The Oral Proficiency Interview (OPI) is "a structured, live conversation between a trained interlocutor/rater and a test-taker on a series of topics of varied language difficulty" (Chalhoub-Deville, 2001). The scoring of the interview is based on the ACTFL Proficiency Guidelines. The Simulated Oral Proficiency Interview (SOPI) was developed by the Center for Applied Linguistics and is a tape-mediated speaking test in which the tape prompts several topics of varied language difficulty. The test takers record their responses onto the cassette tapes and their speeches are rated using the ACTFL Proficiency Guidelines.

¹² The Cambridge Language Assessment Series. See Note # 2.

REFERENCES

- Alderson, J. C. (2000). *Assessing reading*. Cambridge: Cambridge University Press.
- Alderson, J. C., Clapham, C., & Wall, D. (1995). *Language testing construction and evaluation*. Cambridge: Cambridge University Press.
- Alderson, J. C & Hamp-Lyons, L. (1996). TOEFL preparation courses: A study of washback. *Language Testing*, 13 (3), 280-297.
- Alderson, J. C., Nagy, E., & Överges, E. (2000). *English language education in Hungary, Part II: Examining Hungarian learners' achievements in English*. Budapest: The British Council Hungary.
- Alderson, J. C. & Wall, D. (1993). Does washback exist? *Applied Linguistics*, 14 (2), 115-129.
- Anderson, N. J., Bachman, L., Perkins, K., & Cohen, A. (1991). An exploratory study into the construct validity of a reading comprehension test: Triangulation of data sources. *Language Testing*, 8, 41-66.
- Bachman, L. F., Davidson, F., Ryan, K., & Choi, I. C. (1995). *An investigation into the comparability of two tests of English as a Foreign Language*. Cambridge: Cambridge University Press.
- Bachman, L. F. & Palmer, A. S. (1996). *Language testing in practice*. Oxford: Oxford University Press.
- Biber, D., Conrad, S., & Reppen, R. (1998). *Corpus linguistics: Investigating language structure and use*. Cambridge: Cambridge University Press.
- Buck, G. (2001). *Assessing listening*. Cambridge: Cambridge University Press.
- Chalhoub-Deville, M. (2001). Task-based assessments: Characteristics and validity evidence. In Bygate, M., Skehan, P., & Swain, M. (Eds.), *Researching pedagogic tasks: Second language learning, teaching, and testing* (Chapter 10, pp. 210-228). Harlow, England: Longman.
- Douglas, D. (1999). *Assessing languages for specific purposes*. Cambridge: Cambridge University Press.
- Heaton, J. B. (1988). *Writing English language tests*. London: Longman.
- Hughes, A. (1989). *Testing for language teachers*. Cambridge: Cambridge University Press.

- Kennedy, G. (1998). *An introduction to corpus linguistics*. London: Longman.
- McNamara, T. (2001). *Language testing*. Oxford: Oxford University Press.
- Read, J. (2000). *Assessing vocabulary*. Cambridge: Cambridge University Press.
- Ross, S. & Berwick, R. F. (1992). The discourse of accommodation in oral proficiency interviews. *Studies in Second Language Acquisition*, 14 (2), 159-176.
- Weir, C. J. (1993). *Understanding and developing language tests*. New York: Prentice Hall.

An Interview with Dorry M. Kenyon¹

Nathan T. Carr

University of California, Los Angeles

Viphavee Vongpumivitch

University of California, Los Angeles

PROFILE

Dorry M. Kenyon is the Director of the Language Testing Division at the Center for Applied Linguistics in Washington, D.C. and serves as the Book Review Editor for the international journal *Language Testing*. He has done extensive work in test validation studies, particularly in the validation of oral proficiency rating scales (e.g., Kenyon, 1998; Kenyon & Tschirner, 2000; Stansfield & Kenyon, 1993, 1996). He is also well known for his work in oral proficiency testing, such as the Basic English Skills Test (BEST) (Center for Applied Linguistics, n.d.b), Simulated Oral Proficiency Interview (SOPI) (Center for Applied Linguistics, n.d.d), and Computer-Based Oral Proficiency Instrument (COPI) (Center for Applied Linguistics, n.d.c).

INTRODUCTION

This interview with Dr. Kenyon addresses a range of issues based on his experience developing and validating a number of widely used language tests. In the first section, we ask Dr. Kenyon about his education and professional experience, what drew him to the field of applied linguistics, and how he became involved in language testing. In the next section, Dr. Kenyon discusses the Center for Applied Linguistics (CAL) and its past and current research and test development projects. CAL is a private non-profit organization headquartered in Washington, DC, which “aims to promote and improve the teaching and learning of languages, identify and solve problems related to language and culture, and serve as a resource for information about language and culture” (Center for Applied Linguistics, n.d.a). Dr. Kenyon explains the process through which the Center, and specifically the Language Testing Division, takes up research and development projects, and he details several of the projects on which he and his division are currently working.

The discussion in the third section turns to oral proficiency interviews and specifically to the American Council on the Teaching of Foreign Languages (ACTFL) Oral Proficiency Interview (OPI) (Language Testing International, n.d.).

The OPI, which uses a single scale—the ACTFL Language Proficiency Guidelines (Breiner-Sanders, Lowe, Miles, & Swender, 2000)—to rate proficiency in a number of languages, is widely used in the foreign language education community in the United States. In some cases, however, it is either impossible or impractical to have a live rater available to conduct the interview, so institutions sometimes turn to tape- or computer-mediated versions of the OPI. Dr. Kenyon discusses these tests and how they differ from the traditional OPI. We also ask Dr. Kenyon about the ACTFL Guidelines themselves; he addresses such issues as the place of the scale in foreign language education, its origins, the reasons for its popularity, and concerns that have been raised regarding the scale.

In the fourth section, the interview takes up test validation. Validation studies are often underappreciated outside professional testing circles, but they are essential to determining whether a test measures what it purports to measure, and therefore whether its use for a given purpose is justifiable. Dr. Kenyon discusses issues involved in these studies, including real-world constraints on resources and the importance of convincing test users of the need for test validation. In the final section, we ask Dr. Kenyon to address issues he has encountered in the emerging use of computer- and web-based language testing. This is an area of widespread and growing interest in the language testing field, and one in which Dr. Kenyon has extensive experience.

THE INTERVIEW

Professional Background

Carr: *Please describe your academic training and what drew you into the field of applied linguistics and language testing.*

Kenyon: Well, as long as I can remember I've loved language and different languages. Listening to different languages on short-wave radio was a hobby when I was a kid, and trying to identify the different languages. I'd always loved math too, and my goal when I went to college was to be a math teacher, but because I also enjoyed languages so much, I wound up majoring in economics and German, and spent my junior year abroad in Germany.

I learned about a program at the American University in Cairo that offered a master's in teaching English as a foreign language (TEFL). Because I just enjoyed languages and wanted to experience a non-Western one, I decided to do my degree there. I chose TEFL because I was thinking about teaching math versus teaching languages, and math seemed to be more "by yourself," you know. With math, you're thinking things through and working by yourself, while teaching languages is more about working with other people. While working on my master's degree, I taught at the American University at Cairo, and I worked as a teaching assistant at a German high school where I'd help out the English teachers. I also taught four

summers at The American School in Switzerland (TASIS). In all those teaching situations, I was drawn to the testing, especially the placement testing. I always enjoyed that. I'd always be the first to volunteer to score essays and to give oral interviews and things like that.

Coming back to the States after finishing my master's, I taught for several years in the D.C. area, at George Mason University in adjunct positions. I realized that the only people who had stable positions were the directors of English language institutes; the others were all part-timers or paid by the hour and had no benefits or anything. So I said, "Well, I'd better get a Ph.D." The University of Maryland had good PR for its Department of Measurement and Applied Statistics and Evaluation. I had my math background, and they demonstrated that all the people got jobs when they graduated, so I decided to enter that program.

Just at the time that I'd decided to enter the University of Maryland, I met Charles Stansfield, who had come to the Center for Applied Linguistics (CAL), and he hired me to work on a language testing project there. I started my Ph.D. program at Maryland and my work at the Center for Applied Linguistics in the same week, in September of '87. I earned my Ph.D. in educational measurement, but always applied everything to language testing on projects at the Center for Applied Linguistics. It was kind of a natural. To me it was a very satisfying way of using a combination of my math and language backgrounds.

Research and Test Development at the Center for Applied Linguistics (CAL)

Vongpumivitch: *So you're now the Director of the Language Testing Division at CAL, the Center for Applied Linguistics. Would you mind explaining to us what CAL does, what its role is, and also in particular what your division does?*

Kenyon: CAL is a little bit hard to describe in the sense that we do so many things. We have six separate divisions now, and we're involved in all sorts of projects. Probably the most visible one to many would be the ERIC Clearinghouse on Languages and Linguistics, which we have the contract to run, so that's one of the 16 or 17 ERIC Clearinghouses, and we've had that contract for a long time. Maybe others would know NCLE, the National Clearinghouse on Adult ESL Literacy Education. It's now a center, actually, which produces a lot of publications. CAL itself has been around for over 40 years. It's a private non-profit organization and has been involved in all sorts of issues at the intersection of language and society, language and education, and of course one of those issues in language learning is testing. What the Language Testing Division does is work on a whole constellation of projects. We bring in contracts and projects that relate to the assessment of English as a second language, or for example, Americans learning foreign languages.

Vongpumivitch: *So basically the Language Testing Division works on a project when someone contacts you and you sign a contract with them.*

Kenyon: Either as a contract, where they have something specified, or we go after research grants, and that gives us a little bit more leeway to propose what we'd like to do. A lot of our work in the simulated oral proficiency interviews—SOPIs—has come through those grants.

Carr: *When you pursue research grants, are you basically constrained by what someone is already willing to pay for to fund one of your projects, or are you able to decide "I want to look at this. I think CAL needs to explore this area," and then persuade funding bodies to cut a check for it?*

Kenyon: Both. Definitely both. For example, one government agency recently issued a request for the development of a test for a very specific purpose that we applied for. The purpose was to determine, in a machine-scorable way, whether candidates who were taking a translation test out of one language into English could write an English paragraph. But they wanted to have a pre-screen for people, so they wouldn't have to bother administering the hour-long writing test and scoring it. So that was very constrained and they knew exactly what they wanted and it had to be machine scorable. Within that boundary, then, CAL can say, "OK, but let's try it this way," or "Let's look at it this way." I don't know whether we are going to get that project, but I had to propose a different way of conducting it than the outline they sketched out, because the outline they sketched out wouldn't have worked and given them a defensible product at the end. So that's an example of where a request is very prescribed.

An example of the opposite, where you go to a funding source and say "Here's a need that needs to be addressed," was when the Center for Applied Linguistics developed what was called the Basic English Skills Test in the early 1980s for adult ESL. This test is kind of at the survival level, and it's being used more and more nationally for accountability purposes, but that's not what it was intended for at all. So we brought to the attention of the Office of Adult Vocational Education that "Hey, this needs to be updated and needs to be a little tighter for the purposes it's currently being used for." And over several years we were able to secure some funding for that.

Carr: *I gather CAL does a lot of work with government contracts. Is it a common problem with government projects that you create a test for one agency or one purpose and then somebody decides that they could save a little money, and says "Well, it's an English test or a Spanish test, so let's use it over here, too."*

Kenyon: Well, I don't think that anything CAL has developed has been used in that way, but that's often a big issue. In fact, we recently heard that the Office of

Bilingual Education is thinking about national standards for nursing tests for non-English speakers, and they wanted to find out more about the way performance-oriented assessment was developed in Australia in the Occupational English Test. Perhaps you're familiar with it, Tim McNamara (1996) used it to illustrate his book *Measuring Second Language Performance*. They wanted something like that, combined with the TOEIC (Chauncey Group International, n.d.), which is a multiple choice test. And you know, it's really like apples and oranges, and you have to say *so. Somebody will call you up and ask for information and your opinion about using a particular test, and you have to say "Well, what is it, what do you really want?"*. Ultimately, *maybe that's a good thing. I guess you can try to convince them that they should develop something from scratch, or at least adapt something. But there is a lot of that going on. I don't think that has happened with anything that CAL has developed except for the BEST, but that's more because there just aren't many standardized adult ESL assessments out there. The BEST is just about the only oral test that's out there.*

Vongpumivitch: *What are the projects that you're working on at this moment?*

Kenyon: One of the main ones is what we're calling the CBEST, but we may have to change that name because there is a California test called the CBEST (California Commission on Teacher Credentialing & National Evaluation Systems, n.d.). We'll probably call it the Computerized BEST or BEST Plus, because with this new generation of the BEST that we're developing the administration will be assisted by a computer. For example, we're thinking about what questions we'll ask so that we can better get at the target language situation and the ability level of the examinee. That's a big project. Another project is item development for the Foreign Language National Assessment of Educational Progress, which will likely be administered in 2003. We have a new administration now, and their plan for the uses of data from NAEP—that's the National Assessment of Educational Progress (National Center for Education Statistics, n.d.)—is different; they're conceptualizing it a little bit differently.

Carr: *How is it different?*

Kenyon: It's given in a variety of subject areas, and finally after 30 years they've reached foreign language education. But the Bush Administration would like, from what I understand, to have the NAEP serve as a national benchmark against which the states can compare their state-level performances in the basic skills, so they would particularly like it to be given annually in reading and math. NAEP is currently mandated to be administered in grades 4, 8, and 12. If the funds go into doing reading and math on a yearly basis, they won't have funds left over to do this foreign language assessment. But we're moving ahead on that, and we are currently funded.

It's just a case of a limited pot of money, so if we're going to use the money that's there to do annual testing in fourth and eighth grade reading and math, you don't have the other funds left over. Last Saturday, President Bush did his weekly radio address in Spanish, so I don't think it's because the new administration doesn't like foreign language. It's just because conceptually, they see the purpose of NAEP differently. But it will take a while for any changes to take place, because although NAEP is funded by the government, there's a governing board that is independent, and so it has to pass through that. Anyway, we're working on the development of those items. In particular the Center for Applied Linguistics is responsible for developing the interpersonal task for the speaking assessment. And ETS (the Educational Testing Service) has the main grant to do that project.

Another project, which I think you both heard about at LTRC (the Language Testing Research Colloquium)², is the Web test project we're doing (Malone, Carpenter, Winke, Kenyon, 2001), which involves creating a framework for the development of a listening and reading comprehension test that can be scored on, and validly aligned with, the ACTFL Guidelines. Those guidelines have just become really entrenched in the foreign language field, and people like the idea of being able to say "Well, this person can read at the advanced level in Russian, and this person can read at the advanced level in Arabic," and think that we're comparing things that are similar. The importance of this project is that although the government has some ways of testing reading, outside of the government, there haven't been ways of testing listening and reading using the ACTFL scale that have proliferated. So we're developing a framework that could be applied to less commonly taught languages, and developing that in Arabic and Russian for delivery over the Web.

Then another major project that we're working on is something completely different. It's a large-scale project that is funded by the National Institutes of Health (NIH) ultimately, and OERI, the Office of Educational Research and Improvement, and one of the institutes in NIH is the NICHD, the National Institute of Child Health and Human Development. What they've done is they've pooled a lot of money to have a coordinated effort on research to find the best way that children entering our school system speaking Spanish can become literate in English. They funded two five-year programs, and CAL has one jointly with Harvard and Johns Hopkins University. The University of Houston has the other, and then there are other small, independent one-, two-, and three-year projects. They're all coordinated together, so at the end of five years we'll have some definitive research on these issues.

What the Language Testing Division is doing is assisting the project researchers from Harvard and also those from CAL in developing two things. One is developing the instrumentation, because in a lot of this research, at the end of the day sometimes you feel the outcomes were an artifact of poor instrumentation, and so they'd like to have a more principled way of developing instruments. They're not necessarily for language proficiency. Often they're very focused on cross-linguis-

tic issues like morphology, or sound symbol knowledge, or spelling ability as well as more global measures. Another big issue that we're helping this program on is the use of standardized assessments for this population. For example, there's a Woodcock-Johnson test for English speakers (Woodcock & Johnson, n.d.), and there's Woodcock-Muñoz for Spanish speakers (Woodcock & Muñoz, n.d.). But these are bilingual students, so you have to make accommodations using either test. For example, in the test, in order to assess the children's ability to know sound-symbol correspondence, they use pseudo-words. Well, in Spanish, some of those pseudo-Spanish words are actual English words. So if the student pronounces it as a sight word the way it should be said in English because they know the word in English, it would be counted wrong – because they're not showing their knowledge of the sound-symbol correspondence in Spanish. So we have to standardize how those accommodations are going to be made across all the different projects. Those are the main projects I'm currently working on.

Oral Proficiency Interviews and the ACTFL Guidelines

Vongpumivitch: *You also have the Simulated Oral Proficiency Interview (SOPI), which is a tape-mediated Oral Proficiency Interview. What's the difference between the COPI (Computer-Based Oral Proficiency Instrument) and SOPI?*

Kenyon: The SOPI is on tape, so when the tape starts playing, the students get the same exact tasks, the pauses are timed for them, and the response time is timed for them. In the COPI, which is administered over the computer, the tasks are very similar, but the students has more control. They can pick from several tasks. They can say "Oh, this task was too easy, give me one that's harder," or "This task was too hard, give me one that's easier," and they can control the thinking time and the response time. Also, because you can store so much more in the computer than on a tape, they have other choices about the language of instructions. Generally, these are foreign language tests, so the instructions are given in English and the responses are in the foreign language. But higher-level examinees would like to have the instructions in the language that they're studying, so they're not switching back and forth. We couldn't do that on the tape version because you can only put so much on a side of a tape.

Vongpumivitch: *What are the uses of those tests?*

Kenyon: They are mostly developed as research projects, and the COPI specifically was a research format. But they are made available, and we have what we call the self-instructional rater training kits, so people can buy it off the shelf, learn how to rate it, and administer and rate it themselves. They are often used in programs that want to assess the speaking ability of students. For example, a small college might use it because they want to evaluate their majors in French or Ger-

man with a more standardized oral assessment, and in the SOPI we try to relate the outcomes to the ACTFL performance levels so that they get some sense of where their students are in the ACTFL guidelines. So that's the main use. Also, high schools use it for students who have had several years of study, for evaluation purposes, but those are for internal purposes, because they're rating their own.

Carr: *You said the COPI is self-adaptive, more or less?*

Kenyon: Yeah, more or less. There's an underlying algorithm, so that students have to be assessed at tasks at their starting level and also more challenging tasks, whether they like it or not, because you don't want to disadvantage students who really can do more, but self-assess themselves at the beginning at a lower level. So that's really the danger there.

Carr: *Or a student who's too intimidated to try a more challenging level.*

Kenyon: Right, but who may really be able to handle it.

Vongpumivitch: *Does CAL also administer the OPI?*

Kenyon: No, we don't. We do work more and more with ACTFL who provide training for the OPI. Also particularly through their LTI, Language Testing International, they arrange for official OPIs to be given. Often they're given telephonically.

Vongpumivitch: *And the use of OPI is—as opposed to SOPI, which is a research project—I mean the OPI is an actual “test.” Right?*

Kenyon: Through LTI, you can take the test, pay money and get a certificate, and then there's very vigorous quality control. There are only certain people who can do that. ACTFL will also train interested people in how to administer and score the OPI. Those people may go back to their universities and administer the OPI, but again, like the SOPI score, it wouldn't be certified outside of that particular university.

Vongpumivitch: *We'd like to turn to another tool associated with oral proficiency interviews. I think that for a period of time you did a lot of studies validating the ACTFL scale for oral proficiency interviews within many different situations (e.g., Kenyon, 1998; Kenyon & Tschirner, 2000; Stansfield & Kenyon, 1993, 1996). So what would you say are the main findings?*

Kenyon: This is a big issue, and unfortunately, there's a big divide between foreign language education on one side and ESL/applied linguistics on the other. So I

have to say that up front. I have to say up front too that in the foreign language field, the ACTFL Guidelines are quite entrenched. They're not really used much in ESL, EFL, and applied linguistics. One of the reasons is because clients want things interpreted on the ACTFL scale. It's just a given in foreign language education, a little bit like what's developing in Europe with the Council of Europe's (1996) six levels, with three main levels, each with two sublevels. CAL has sought to develop products that meet the needs that people have and so we've been working with the ACTFL scale with the SOPI for a long, long time. And, as we've had opportunity, we've tried to look at some of the issues related to the use of the scale. We can't get funding for a project that includes validating the whole ACTFL scale, but as we're doing a project we can look at validating different pieces of the scale.

The main published research that you might be thinking about is on students' perceptions of difficulty in the SOPI. In the OPI, the questions are targeted at these different main levels. And one issue is, are those tasks really more difficult? Finding the difficulty of tasks is a big problem, a big issue, and how do you really define the issue? There are so many things that make a task more or less difficult. So, some of that research (e.g., Kenyon 1998) was taking tasks like the ones used in the SOPI which are developed to assess at these different ACTFL levels, and seeing if students put them in the same kind of difficulty order as would be predicted by the scale. There are several studies of that through a self-assessment instrument that we've done. Of course, the scale and this all come from the government beginning in the 1950s.

Carr: The ILR (Interagency Language Roundtable)?

Kenyon: The ILR skill level descriptors. This is also catching on a whole lot in business too, because they're the ones that ACTFL mostly does language testing for. So it's just become a common standard out there. I think a big concern, both in the earlier days and still today, is inter-rater reliability. People who are using this test seem to be focused a lot on that kind of research. So as I say, we've been trying to look at it — as we've had opportunity — in different ways. My sense of it, when it all boils down, is that in broad brush strokes it has some use for painting things broadly, but it's not the end-all and be-all and I think that some people, proponents especially in the early days, were saying that it was. I think better describing what it is and what it isn't is probably the most honest thing you can do with it. Because it's there to stay, so people should understand the nature of the beast, so to speak.

Vongpumivitch: *I can understand why it's so popular, because when you need to make a quick decision, to make a quick judgement on someone's proficiency level, those scales come in so handy. Apart from those scales, is there any other kind of scale that is that clear and "comfortable?"*

Kenyon: I think you've said something that's really very important. It's that it's the practitioners in foreign language education who like it for that reason. Because I think it responds to something in their experience and in broad-brush strokes. Let's take intermediate-mid, because that's a very broad level. Students who are intermediate-mid level have many different profiles, and if you did diagnostic testing, individual strengths and weaknesses would be very different. But in a broad brushstroke, people find it useful just to call them intermediate-mid. However, I digressed a little bit. The acceptability of a scale is a bit of a social phenomenon, too. Researchers could say many things about that scale, or about something else. Like, you know, the TOEFL (Educational Testing Service, n.d.b) has the TOEFL 2000 Project, which is supposed to respond to some of its weaknesses in the past, and I applaud ETS for doing that. But the TOEFL is pretty entrenched. Researchers can critique it, but if the schools and colleges and universities use it and find that it's doing its job, or it's doing a job, or it's helpful, that, in a sense, can have more strength than what the researchers say. I really applaud ETS for trying to bring the two together. I know that they've been putting a lot of resources into that and I expect to see really good things in the future. But users are what make it entrenched, and in the United States there aren't any other scales that have been adopted. In adult ESL education there's something called student performance levels, or SPLs, in the MELT (Mainstream English Language Training) Project, which gave rise to the BEST, and they have a somewhat similar currency in adult ESL. Most of the adult ESL tests try to map onto some of the SPLs. I can't think of another scale, with the exception of the one in Australia³ (Wylie & Ingram, 2001) which I understand as being derivative from ILR and ACTFL. Then there also are the Council of Europe proficiency levels.

Carr: *Some have criticized the ILR and ACTFL scales, though, because they contain both content and context as parts of the definition of proficiency. Specifically, they measure the ability to use vocabulary and grammatical structures in the context and under the conditions that are included in the testing procedure. You've worked extensively on projects using these scales, and they do have a lot of currency. What do you think would be the response from the ACTFL side?*

Kenyon: I don't know what to respond from the ACTFL side. I can say how I'm looking at it right now. I think proficiency is much, much narrower than communicative competence. It's kind of necessary—maybe not always necessary, I mean, people communicate in the real world without knowing each others' languages, but that can only go so far. I'd say it's necessary but not sufficient to accomplish real-world tasks. I think when all these things started in the government, it needed to develop a scale to know what the people who it was training for embassies could do in the real world. So they—and this was the 1950s, too—were trying to replicate the type of language in the interview context that might be found outside the interview context. Since that time, I think we've become sensitive to the many

other factors that are involved in being successful in a communicative situation, and I think what you were talking about was content?

Carr: *And the specifics, and problems with generalizability...*

Kenyon: I think in those early days they overgeneralized and some claims were made that were very, very broad. They were probably hard to support because having a face-to-face interview with one person, well, how will you do outside that context? On the other hand, if you don't have that proficiency, that's necessary but not sufficient, that's what you get out of an OPI, I think, in terms of what's been automatized, what can they speak freely about, what are the content areas, what's the breadth of their grammatical control and their vocabulary and the way that they can express themselves. That's going to be necessary out there, but it clearly isn't sufficient. It's not everything that's needed out there in the world. I would agree with you that the claims for generalizability outside of the context are probably exaggerated in that sense.

Vongpumivitch: *You mentioned raters, and that in using the scale-based test, raters are a very important group of people, and that a lot of research has been done on rater training. Who are the raters for these tests, such as the OPI?*

Kenyon: For the OPI, again, there's the official LTI OPI, and those are people who work very closely, and they're rating almost daily now. I mean, the volume of work has gone up exponentially. So there's tight quality control on those people, and they're meeting, they're benchmarked, there's all sorts of double-rating and triple-rating for an official ACTFL OPI through the LTI. The unofficial ones might be analogous to the SPEAK (ETS, n.d.a), you know, to what you do here at UCLA. Hopefully you have competent people there who have been trained who are making these ratings. Maybe there's a calibration test and everything like that.

Test Validation Studies

Vongpumivitch: *As someone who has a lot of experience doing validation studies, either for rating scales or for tests, what is the goal, in your opinion, of a validation study, either for a test or for a rating scale, and what is the best way to go about doing a good validation study?*

Kenyon: The rating scale in and of itself isn't validated apart from the whole assessment process. That's just one component there, but as you know, the standard issue of validity is what are the inferences being made, what are the actions being taken about the student on the basis of this test, and what theoretical considerations along with empirical evidence can we use to demonstrate that it's appropriate to make those inferences. Essentially, the question is: How do you justify

the use that's being made of these test results? So how do you do that? Well, first you have to have a very clear understanding of how the test results are being used. And I think that's one thing that may not be clear to everybody about the ACTFL guidelines, because they're used in so many different contexts. For example, with the LTI, now it's often used for correct placement in employment positions, so in that context you'd have to understand the decision that's going to be made about that candidate for that position.

Let's give the example of someone who's going to be employed with the AT&T Language Line, a service you can dial up for translation in one of 140 different languages. Anybody can do that on line. So if you're a border policeman and you stop somebody and you can't understand what language it is, you call up this Language Line. I don't know whether AT&T still runs it, but the Language Line service is still out there. You call them up, get the language identified, and they get somebody who can translate on the phone while you're trying to question this person that you stopped at the border. So who fills those positions? Well, people who fill those have to know English, and they have to know the language that's being spoken. How well do they have to know it? Well, let's say they say that ACTFL "superior" is what's needed. So the company that runs that business has to hire some people in Thai and say that these are "superior" level people. So they give the OPI in Thai and find out whether these people are "superior" or not.

Well, the question is: In validating that, is being "superior" in Thai, and being "superior" in, let's assume English, sufficient to do that job or are there other skills necessary? I think there might be other skills involved in interpreting that might be necessary to train on, but if you don't have those language skills, it may not matter much. So if an organization is hiring people and putting them in this job just on the basis of an OPI, that might be insufficient because there might be other skills involved. So that might be a validity issue. If the OPI were to say "All you need to do is have an OPI and you can do this job." Well no, there's probably more. Again, going back to that sense of being necessary but not everything. Validation is so contextual; it's hard to say there's one way to validate. If the organization were to say, "Well, we can't provide training, we can only make decisions based on the OPI," and the people from the OPI are saying, "That's fine, sure, this is sufficient," well that might be problematic.

Vongpumivitch: *As graduate students, we can take classes about validation studies, and read about validation studies, but I think we really need to hear the views of people who have actually done validation studies.*

Kenyon: Well, the issue – and this is serious – is that in academia when you're not working with it, you think there are unlimited resources. Resources are very, very limited, and validation is the last step. It is easy to use all your money up before you get to that. I think one thing that language testers have to remind projects from the very beginning is that, at the end of the day, you're going to have to provide

demonstration for the uses that are being made of the assessment results. We have to set aside funding for that. I think that really is important, but often that can be overlooked in the real world. One of my pet peeves with working with people in the real world is they often start with what the task looks like. They don't think through the issues. And that's why my workshop at this conference is very different from what was in the abstract. I changed it. It's from thinking through those issues and then thinking about what the task is going to look like. But people say, "Oh, this TOEIC test looks like a good one. Now we need one for Spanish. Let's just translate it into Spanish, you know. But we need it for a whole different purpose. We want it for teachers." when it was written for business contexts. They'll look at something, at the superficial form, and say "That's what I want, I just want you to revise it. It's not going to take very much to revise it."

Computer-Based and Web-Based Language Testing

Carr: *You already talked to us about the ways that the COPI differs from the SOPI. In general, what would you see as some of the advantages and disadvantages of computer-based testing, web-based testing? What are the or doors that it opens and limitations that it imposes?*

Kenyon: That's a good issue for us in our situation. Again, given that we have limited resources, there seems to be tons of potential out there for what could be done. But when you get into real-world projects, often people will fund the tried and true. So basic research in what could be done and really exploited to make computer-based tests more than just paper-based tests on computer is really necessary and valuable. But my general sense of it is that large-scale programming in language testing, at least, computer-based testing, especially computer adaptive testing, has not been giving the return on investment, so to speak, that was originally foreseen and desired and hoped for. I'm aware of some computer adaptive tests that have run into issues and problems.

Carr: *What kind of issues have you seen occurring?*

Kenyon: The big issue is with the development of the number of items. There's usually an incredibly large item bank that's necessary to support a computer adaptive test. So that's one big issue, getting that number of high-quality items, having them all calibrated somehow, having them all field tested, revised, and calibrated before going into the pool. It's an expensive and big task for smaller assessment programs, and even for big ones.

CONCLUSION

Dr. Kenyon raises a number of important issues regarding language testing. Perhaps most significant are his comments relating to the use of the ACTFL Guidelines and OPI in the foreign language education community and some of the suggestions he provides on how to promote test validity and validation in real-world projects.

Regarding the ACTFL Guidelines and the OPI and its "cousins," he argues *that in spite of issues involving generalizability of performance beyond the context of the interview, these tests are widely popular in the foreign language community and unlikely to be displaced in the foreseeable future. This may be in part a symptom of a wider problem: In describing the entrenched position occupied by the ACTFL Guidelines, Dr. Kenyon mentions a disconnect in the United States "between foreign language education on one side and ESL/applied linguistics on the other." This is a regrettable and somewhat disturbing trend, as such a gap is potentially harmful to both communities, posing the risk of cutting off foreign language education professionals from research in applied linguistics, and of limiting the opportunities open to applied linguists in general and language testers in particular for doing research in languages other than English.*

Finally, Dr. Kenyon notes that in the real world resources are limited, making it important to point out from the beginning to test users that funding must be set aside to pay for validation in order to support the uses that will be made of test results. In addition to recommending the development of a new test or adaptation of an appropriate existing test, when asked for advice on the potential adoption of an inappropriate assessment procedure, Dr. Kenyon proposes another way in which language testing professionals can help encourage more valid test use. The other way, which might be termed a "half a loaf" approach, is somewhat more subtle, but well worth noting: When working on a project involving a portion of a larger testing program, opportunities should be found to do limited validation studies of some specific aspect of the test. While it is obviously preferable for test users to have invested in comprehensively investigating the validity of a test's uses, when they have failed to do so, language testers may be able to use this approach to at least partially correct matters.

NOTES

¹ The interview took place at the Fourth Annual Conference of the Southern California Association for Language Assessment Research held in Pasadena, CA (May, 2001). Nathan Carr and Vipavee Vongpumivitch are Ph.D. students in applied linguistics at the University of California, Los Angeles, and are specializing in language assessment.

² The 23rd Annual Language Testing Research Colloquium (LTRC) was held in St. Louis, MO in February, 2001.

³ The International Second Language Proficiency Rating Scale (ISLPR), formerly the Australian Second Language Proficiency Rating Scale (ASLPR), which is "widely used in Australia to assess the general language proficiency of adult ESLK learners" (Brindley, 1995, p. 3).

REFERENCES

- Breiner-Sanders, K. E., Lowe, P., Miles, J., & Swender, E. (2000). ACTFL Proficiency Guidelines—Speaking, Revised 1999. *Foreign Language Annals*, 33, 13-18.
- Brindley, G. (1995). Introduction. In G. Brindley, (Ed.), *Language assessment in action* (pp. 1-9). Sydney, Australia: National Centre for English Language Teaching and Research.
- California Commission on Teacher Credentialing & National Evaluation Systems (n.d.). *California basic educational skills test (CBEST)*. Sacramento, CA: Authors.
- Center for Applied Linguistics (n.d.a). *About CAL*. Available: <http://www.cal.org/admin/about.html>
- Center for Applied Linguistics (n.d.b). *Basic English skills test (BEST)*. Washington, DC: Author.
- Center for Applied Linguistics (n.d.c). *Computerized oral proficiency interview (COPI)*. Washington, DC: Author.
- Center for Applied Linguistics (n.d.d). *Simulated oral proficiency interview (SOPI)*. Washington, DC: Author.
- Chauncey Group International (n.d.). *Test of English for international communication (TOEIC)*. Princeton, NJ: Author.
- Council of Europe (1996). *Modern languages: Learning, teaching, assessment. A common European frame of reference*. Available: <http://culture.coe.fr/lang/eng/eedu2.4.html>.
- Educational Testing Service (n.d.a). *Speaking proficiency English assessment kit (SPEAK)*. Princeton, NJ: Author.
- Educational Testing Service (n.d.b). *Test of English as a foreign language (TOEFL)*. Princeton, NJ: Author.
- Kenyon, D. M. (1998). An investigation of the validity of task demands on performance-based tests of oral proficiency. In A. J. Kunnan (Ed.), *Validation in language testing: Selected papers from the 17th Language Testing Research Colloquium, Long Beach* (pp. 19-40). Mahwah, NJ: Lawrence Erlbaum Associates.
- Kenyon, D. M., & Tschirmer, E. (2000). The rating of direct and semi-direct Oral Proficiency Interviews: Comparing performance at lower proficiency levels. *Modern Language Journal*, 84, 85-101.
- Language Testing International (n.d.). *American Council on the Teaching of Foreign Languages oral proficiency interview (ACTFL OPI)*. White Plains, NY: Author.
- Malone, M., Carpenter, H., Winke, P., & Kenyon, D. (2001, February). Development of a web-based listening and reading test for less commonly taught languages. Work in progress session presented at the Language Testing Research Colloquium, St. Louis, MO.
- McNamara, T. F. (1996). *Measuring second language performance*. London and New York: Longman.
- National Center for Education Statistics (n.d.). *National assessment of educational progress (NAEP)*. Washington, DC: Author.
- Stansfield, C. W., & Kenyon, D. M. (1993). Development and validation of the Hausa Speaking Test with the "ACTFL Proficiency Guidelines." *Issues in Applied Linguistics*, 4, 5-31.

- Stansfield, C. W., & Kenyon, D. M. (1996). Comparing the scaling of speaking tasks by language teachers and by the ACTFL Guidelines. In A. Cumming & R. Berwick (Eds.), *Validation in language testing* (pp. 124-153). Clevedon, UK: Multilingual Matters.
- Woodcock, R., & Muñoz, A. F. (n.d.). *Batería Woodcock-Muñoz-Revisada (Batería-R)*. Itasca, IL: Riverside Publishing.
- Woodcock, R., & Johnson, M. B. (n.d.). *Woodcock-Johnson III Complete Battery*. Itasca, IL: Riverside Publishing.
- Wylie, E., & Ingram, D. (2001). *International second language proficiency ratings (formerly the Australian second language proficiency ratings): About the ISLPR*. Available: <http://www.gu.edu.au/centre/call/content4.html>.

Language Testing by Tim McNamara. Oxford: Oxford University Press, 2000, xv+140 pp.

Reviewed by Kirby J. Cook
Teachers College, Columbia University

Language testing has evolved over the past 40 years with fresh approaches, innovative methods, and social policy all contributing to its growth. Though these issues are fundamental to the language tester or researcher, many teachers, school administrators, and graduate students not focusing on language assessment have not been exposed to this evolution; very few books cater to the novice tester. Published works have continued to become denser and more technical, often requiring background knowledge beyond those with only a budding interest in the field. In *Language Testing*, Tim McNamara makes the field of language assessment accessible to the untrained reader through logical explanations and a clear organization in a comprehensive, albeit basic, overview of language testing.

Language Testing is part of a seven-book introductory series on language study from Oxford University Press, all of which are organized in four sections: Survey, Readings, References, and Glossary. The Survey, presented in eight chapters, represents the main body of the book. It provides the reader with a comprehensive overview of all major concepts in language testing, including test design, rating procedures, validity, types of measurement, and the social implications associated with assessment. McNamara provides theories from influential literature on language testing, including the works of Robert Lado, John Oller, Dell Hymes, and Lyle Bachman and Adrian Palmer, among many others. He also outlines directions for future research as well as dilemmas within the field.

Chapter 1, entitled "Testing, testing... What is a language test?," provides an opening for this introductory book. Although a bit redundant, the introduction helps to orient the reader and demonstrates the boundaries of the book's content. In this chapter, McNamara appeals to the sensibilities of those *using* tests (or the information they provide) in addition to the test creation experts. He provides useful, familiar, and accessible examples of testing concepts and specific terms in boldface type. Readers may access concise definitions of these terms in the Glossary near the end of the book.

Chapter 2, "Communication and the design of language tests," involves an introduction to communication and its relationship to test design. McNamara brings to attention such terms as *test construct*, *discrete point testing*, *integrative and pragmatic tests*, *cloze test*, and *job analysis*. All of these concepts are carefully framed by the contributions of influential language testing researchers from a historical perspective. McNamara traces the response of the field of language testing

from Lado's skills testing approach to a need for integrative tests and later to pragmatic tests. He briefly introduces some early theories of communicative competence, followed by a discussion of more recent developments in communicative language tests and models of communicative ability.

Chapter 3, "The testing cycle," includes an interesting description of test content and method and provides a cursory introduction to authenticity, response formats, and test specifications. In his discussion, McNamara presents the reader with a basic explanation of the design stage, the construction stage, and the try-out stage. By means of a skillful overview, McNamara manages to capture the essence and evolution of the test design process.

In Chapter 4, "The rating process," McNamara inclusively outlines criteria for rater-mediated assessment. He discusses intricacies and inconsistencies associated with the rating process and includes methods researchers have developed to combat these difficulties. The section dealing with rating scales is perhaps the most useful for the untrained reader; McNamara provides unambiguous explanations for both holistic and analytic rating scales and includes a sample rating scale for exemplification. In alluding to the importance of rating scales and consistency as central to any testing procedure, McNamara prepares the reader for the treatment of test validation in the next chapter.

In Chapter 5, "Validity: Testing the test," McNamara discusses the significance of validation with the same depth as any more advanced text, so the novice reader will understand the importance of validity to the same degree as would any experienced tester. McNamara describes face validity as well as content validity and addresses threats to validity, referring back to test content, test method, and test construct in his previous discussions. He also briefly treats consequential validity (how the implementation of a test may affect the integrity of inferences made about the test-takers), calling into question the social impact of tests.

Chapter 6, "Measurement," contextualizes dense concepts and technical terms from a traditional yet enlightened perspective. Although McNamara uses terms such as *norm-referenced* and *criterion-referenced* measurement, *reliability coefficient*, *normal distribution*, and *item discrimination*, he avoids the use of highly scientific language in his explanations of these concepts. He also discusses tools used to analyze test performance, including Item Response Theory and Rasch Measurement, and touches on more recent developments in computer adaptive testing.

In Chapter 7, "The social character of language tests," McNamara addresses assessment as it relates to social and educational policy, accountability, washback, and test impact. McNamara stresses the responsibility of the testing community in developing fair and ethical tests and presents an informed perspective on the more radical field of critical language testing, which advocates the entire reconstruction of testing due to its relationship to socio-political power.

Chapter 8, "New directions—and dilemmas?," outlines the role of technology in language testing, including the difficulties encountered by those participat-

ing in all aspects of assessment. He provides an overview of computer based testing (CBT) and semi-direct tests of speaking, including issues of validity, authenticity, and responsibility surrounding this new technology. McNamara asserts that "a language test is only as good as the theory of language upon which it is based" (p. 86) and suggests looking to the nature of language and communication to address test vulnerabilities.

The eight-chapter Survey gives the novice reader a helpful foundation; however, a more experienced reader may find this section elementary. Although the introduction in Chapter 1 gives the illusion of adequate topic coverage, some of the explanations, though clear, are too simplistic; the novice reader comes away with a perspective on language testing that is deficient in some compulsory details. Much of the superficial treatment of information contained in the introduction could have been absorbed into the seven more detailed chapters that follow in the Survey. While the chapter content is necessary for the untrained reader, perhaps most useful to the advanced reader are the chapter conclusions, which provide an accurate, succinct, and complete representation of the content coverage.

After the Survey, the second section of the book, Readings, provides additional texts for those readers interested in pursuing in greater detail any of the chapter topics covered in the Survey section. McNamara lists seminal texts according to their corresponding Survey chapter and includes a short summary and questions designed to "give readers an initial familiarity with the more specialist idiom of the linguistics literature, where the issue might not be so readily accessible, and to encourage them into close critical reading" (p. xiii). This section aids in reinforcing the Survey content for the novice reader and requires thoughtful pondering from the more advanced reader seeking comparisons across texts and within issues. Though perhaps the most useful section of the book, the short summaries in the Readings section should not be used as a substitute for reading these primary sources in their entirety.

The third section, References, includes a selection of books and articles for further reading. This annotated list of references includes a tertiary categorization system indicating the level of reading for the interested reader. Novice readers may choose to broaden their knowledge base with other introductory texts, graduate students of assessment can challenge themselves with more advanced or technical reading, and the language testing expert may choose highly specialized and demanding reading. McNamara presents this section as useful for all levels of study in language testing by choosing books of varying levels in almost all topic areas. In this effective reference section, readers will find the longer annotations concise and informative.

The Glossary is the final section of the book and contains definitions of the boldfaced terms from the Survey section. While the terms are fairly well defined, they are cross referenced only with their location in the Survey introduction, where their explanations are basic and decontextualized. Although these definitions may prove helpful to the untrained reader, a more advanced investigator will not find

the Glossary particularly useful. The elaborated, detailed, and contextualized descriptions of the integral terms contained in the Survey chapters will be of interest to all readers, especially more advanced ones. It would have been helpful if the author had included an additional index referencing these terms in the Survey section for those looking for more than a minimal definition.

Despite some of the inherent difficulties traditionally associated with producing an introductory text that is both inclusive and comprehensible, *Language Testing* represents an extremely accessible overview in the abstract and technical field of language assessment. McNamara's introductory perspective captures the fundamental concepts of language testing for the novice reader and provides a comprehensive reference text for the more advanced reader. The Readings section is indispensable for the language assessment learner, and the References section is a valuable contribution for anyone interested in pursuing the field of language testing in any capacity. Although this text is not meant for intensive study, it is a springboard for further study or research. *Language Testing* is an elemental introduction to assessment and should be a part of the library of every individual interested in the field.

How Children Learn the Meanings of Words by Paul Bloom.
Cambridge, Massachusetts: MIT Press, 2000, xii+300 pp.

Reviewed by Masahiko Minami
San Francisco State University

How do children learn the meanings of words? In his new book, Paul Bloom examines a variety of issues associated with children's word learning, a process intricately connected with other aspects of language acquisition. Bloom claims that children learn words via cognitive abilities that already exist for other purposes, such as the ability to infer others' intentions, the ability to acquire concepts, and an appreciation of syntactic structure. Bloom's book provides a series of elegant and convincing arguments concerning how children learn words.

In Chapter 1, "First Words," Bloom lays out the plan for the book and briefly describes issues surrounding the overall topic. In Chapter 2 the author explores *fast mapping*, in which children make a quick guess about a word's denotation on the basis of limited experience. Chapter 3, "Theory of Mind," deals with a wide range of topics, including the listener's ability to determine the references made by his or her interlocutor's choice of words; here also, Bloom investigates children's appreciation of the mental states of others, through which children acquire lexical items (and syntax as well) by means of associative learning. Because the majority of words that children initially acquire are nouns, Bloom gives special treatment to nouns and pronouns: Common nouns are discussed in Chapter 4, and pronouns and proper names are dealt with in Chapter 5.

In Chapter 6, "Concepts and Categories," Bloom extends his analysis to the conceptual foundations of word learning. In Chapter 7, "Naming Representations," he discusses a case study important to any theory of concepts and naming—visual representations. From here, Bloom moves to other parts of speech: In Chapter 8, "Learning Words through Linguistic Context," he offers an account of how children learn verbs and adjectives, as the development of syntactic abilities cannot be dissociated from the development of lexical abilities. Chapter 9 deals with how we learn the words for numbers and Chapter 10 with how the words we learn affect our mental life. In Chapter 11, "Final Words," Bloom provides a brief summary and some general remarks. Throughout the book, the author weaves in ideas proposed by such linguists, psychologists, and philosophers as B. F. Skinner, Noam Chomsky, and Jean Piaget, who, through different lenses, have closely observed and analyzed how human beings develop and how they conceptualize the world around them.

As with most language acquisition texts, Bloom makes early reference to an issue long relevant to human development: the nature/nurture debate. These alter-

natives have been particularly salient in the study of language acquisition. According to Skinner (1957), language learning is based on experience. Skinner's empiricist view stands in striking contrast to Chomsky's (1959, 1965) nativist view that humans have a biological endowment which enables us to discover the framework of principles and elements common to human languages. Bloom argues throughout his book, however, that as more details of language development continue to be revealed, it appears even more improbable that either nativism or empiricism alone could account for the complexity of the language acquisition process.

Chomsky claims that speech is a central aspect of human linguistic ability and that there *are* universal features in language development. Significant similarities have been identified in the patterns of language acquisition across different languages, and this fact suggests that species-specific biological factors play a crucial role in the ability of humans to acquire and process language. Regardless of the language they speak, children first acquire the names of basic objects (e.g., *dog*) and then develop gradually in both the more abstract (e.g., *animal*) and more concrete directions (e.g., *collie*). Generally speaking, therefore, early vocabulary is concrete, and increased abstraction and specification follow.

Yet Bloom asserts that we cannot ignore the influence of environmental factors in language development, referring to positive correlations in vocabulary sizes between parents and children (Chapter 2). Also, Bloom acknowledges the existence of crosslinguistic or crosscultural differences (Chapter 3) in patterns of language acquisition. Because mothers provide linguistic modeling in a variety of ways, mothers are crucial agents of language socialization, and patterns of language socialization may differ across cultures. Bloom mentions, for instance, the Kaluli culture of Papua New Guinea, in which very little sustained dyadic verbal exchange takes place between mother and child; thus, adults appear to give little effort to teaching the meanings of words to their youngsters. In Western societies, mothers generally modify their language to some extent when interacting with their children, and these modifications would seem to accelerate the rate of children's language acquisition. In this way, a great degree of cross cultural variation seems to exist, attesting to the complexity of environmental factors.

Bloom adds in Chapter 4 that while in certain cultures adults do not usually label objects for children, they might nonetheless teach proper names to their children. Thus some type of parental encouragement seems ubiquitous, which might explain why children of other cultures do not necessarily learn vocabulary more slowly. The evidence provided in Chapters 2, 3, and 4 therefore leads to the reasonable inference that heredity draws the blueprint for development, but environmental factors, such as parental encouragement, affect children's word learning. In this regard, in his account of how children grasp the meanings of words, Bloom claims that environment and inborn capacity, as well as maturation, interact in a complex fashion during language acquisition in young children. Even if acquiring nouns early is in part a genetically transmitted trait, the acquisition of even simple

nouns requires a great many conceptual, social and linguistic capabilities that interact in intricate ways.

Bloom further adds that children's word learning must occur in pragmatic contexts in order to provide material for communication. Because the basic function of language is communication, Bloom claims "the best way to learn a word through context is by hearing it used in a conversation with another person" (p. 192). Bloom thus suggests that we should examine the controversy surrounding the nature/nurture debate from a more comprehensive perspective. Clearly Bloom's position does not differ significantly from the emergentist paradigm (MacWhinney, 1999), which claims that domain-general cognitive mechanisms (e.g., working memory) work on environmental stimuli to result in the complex and elegant structures that characterize language. In this sense, Bloom's view is a constructivist one, stressing the interaction between the organism and the environment (i.e., children gradually learn words by interacting with environmental factors such as parents' speech patterns).

An intriguing concept—both cognitively and crosslinguistically—is Bloom's account of the nature of our numerical ability. In Chapter 9 Bloom refers to Wynn (1992), who provides evidence that 5-month-old infants seem able to deal with the quantification of small numbers of objects. In one exemplary experiment, babies were briefly shown a doll that was then hidden behind a screen. The babies then saw a hand place another doll behind the screen, and subsequently the screen was pulled away to show one or three dolls. In response, the babies appeared surprised at the unexpected "wrong" answers, and this is a finding that, though Bloom does not refer in detail to Piaget, clearly challenges Piaget's (1952) cognitive theory that claims that infants cannot think about objects that are not physically present or about past events.

Furthermore, since Bloom emphasizes the existence of some association between the development of theory of mind and the onset of word learning (Chapter 2), his argument in this book inevitably questions the very nature of Piaget's view that thought processes change via a series of stages. According to Piaget (1932), children's speech in the preoperational period (2 to 7 years of age) often lacks the ability to take perspective; their speech reflects the assumption that other people share their own view of things. As Bloom argues, however, if the ability to infer others' intentions is critical for children's learning of the meanings of words, two-year-old children should be able to take another individual's perspective. This argument contradicts Piaget's conclusions about preoperational children's limitations in perspective taking. Because Piaget's substantial underestimation of the abilities of young children is already well known, if Bloom had developed his argument explicitly against the Piagetian framework his ideas would have been even more thought-provoking.

In short, while I feel that Bloom's discussions are unconvincing in some respects and that many of the questions posed remain unanswered, this book is a compelling account of how children learn the meanings of words. Because the

book is written in a clear, engaging, and casual style, it might be accessible even to those who have little background in psychology or linguistics. At the same time, the evidence presented in the book, including abundant and trustworthy data from many independent studies as well, will be equally valuable to those researchers who are engaged in language and cognition studies.

REFERENCES

- Chomsky, N. (1959). Review of *Verbal behavior* by B. F. Skinner. *Language*, 35(1), 26-58.
- Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.
- MacWhinney, B. (Ed.). (1999). *The emergence of language*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Piaget, J. (1932). *The moral judgment of the child*. New York: Harcourt Brace.
- Piaget, J. (1952). *The origins of intelligence in children* (M. Cook, Trans.). New York: International Universities Press. (Original work published 1936.)
- Skinner, B. F. (1957). *Verbal behavior*. Englewood Cliffs, NJ: Prentice-Hall.
- Wynn, K. (1992). Addition and subtraction by human infants. *Nature*, 358, 749-750.

Memory: From Mind to Molecules by Larry Squire and Eric Kandel. New York: Scientific American Library, 1999, 235 pp.

Reviewed by Sheila Crowell
University of California, Los Angeles

Memory: From Mind to Molecules represents the skillful attempt by Larry Squire and Eric Kandel to explain the cellular processes involved in the formation and consolidation of memories. Clearly written and masterfully presented, *Memory* guides the reader through complicated material step by step, such that even non-scientific readers will find the journey worth their while; those with advanced scientific backgrounds, however, will not be disappointed. As one of the first attempts to link research in the behavioral sciences with research in neurobiology, *Memory* highlights issues of interest to any reader who desires to understand the inner workings of the mind. And in the field of applied linguistics, understanding the neurobiological processes involved in learning and memory has become increasingly important.

The questions addressed by Squire and Kandel are presented in the first chapter (p.3): What is memory? How does it work? Are there different kinds of memory? Where in the brain do we learn? Where do we store what is learned as memory? Can memory storage be resolved at the level of individual nerve cells? If so, what is the nature of the molecules that underlie the various processes of memory storage? These questions shape and organize the book, with Chapters 2 and 3 dedicated to the neural architecture involved in nondeclarative and short-term memory, Chapters 4 through 7 explaining the molecular basis for declarative and long-term memory, and Chapters 8 through 10 presenting the ways in which these memory processes influence our lives on a daily basis. Throughout the book, Squire and Kandel elucidate complicated biological phenomena through stories, diagrams, and pictures of artistic masterpieces. These colorful pictures serve to remind the reader of the powerful connection between who we are and how the brain functions. In essence, the entire book is a statement of the inseparability of the fields of biology, psychology, and philosophy.

Chapter 2, which addresses the cellular mechanisms involved in nondeclarative memory, provides the first detailed look at how memory storage takes place at the level of individual nerve cells. Nondeclarative memory is the type of memory that is typically "not accessible to the conscious mind" (p.24), such as memory for motor skills (i.e., riding a bike) or habituating to a non-aversive stimulus. Though nondeclarative memory occurs subconsciously, there are clear changes at the level of individual synapses that occur as a result of learning. For example, in the first neuron to fire (the presynaptic cell) there can be an in-

crease in the number of synaptic vesicles that carry neurotransmitters used in communicating with the second neuron (the postsynaptic cell). With a higher number of vesicles, the probability that the first cell will cause the second cell to fire is much higher, meaning that performing a task will be much easier. Another change in neurons that can result from learning occurs at the structural level. In the case of habituation to a stimulus, a connection between the presynaptic cell and the postsynaptic cell might become completely ineffective, causing the processes on the presynaptic neuron to retract. This decrease in efficacy is the result of an increase in learning, for example, in a case where a person habituates, or becomes accustomed, to a loud noise or bright light not followed by any negative consequences. These basic learning mechanisms operate in accordance with the synaptic plasticity hypothesis, which states that "the ease with which an action potential in one cell excites (or inhibits) its target cell is not fixed but is plastic and modifiable" (p. 35). In other words, though most neurons are wired in a specific way at birth, the strength of the connections between these neurons can change as the result of learning.

Later chapters dealing with declarative memory present further biological and psychological descriptions of the underlying processes involved in the storage of learned experiences. Declarative memory is "memory for events, facts, words, faces, [and] music...[it is] knowledge that can potentially be *declared*, that is, brought to mind as a verbal proposition or as a mental image" (pp. 70-71). Chapters 4 and 5 set up Chapters 6 and 7, with Chapter 4 addressing declarative memory in terms of research in cognitive psychology and Chapter 5 introducing the brain systems involved in declarative memory. The authors' emphasis becomes clear in Chapters 6 and 7, where the discussion is centered on the mechanisms of synaptic storage used for declarative memory and for converting learned experience into long-term memory. In these chapters Squire and Kandel discuss one of the most important and interesting phenomena in the study of learning and memory. This phenomenon, known as *long-term potentiation* (LTP), is now widely believed to be the major cellular mechanism for the storage of long-term memories.

LTP is an increase in synaptic strength that results from high-frequency stimulation to the presynaptic neuron. This strengthened connection between two neurons can last for hours, days, or even weeks (p.111) and involves a number of physiological changes at the cellular level. One important discovery regarding LTP is that for LTP to occur there must be changes in both the presynaptic cell and the postsynaptic cell. These changes include an increase in the release of neurotransmitters from the presynaptic neuron and a depolarization—a reduction in the membrane potential of the postsynaptic neuron.

When this happens, magnesium ions (Mg^{2+}) that are blocking the receptors on the postsynaptic cell are expelled, opening the receptors for an influx of calcium (Ca^{2+}). The calcium ions cause a number of chemical changes that produce a byproduct which diffuses out of the postsynaptic cell, across the gap, or cleft, between the synapses and into the presynaptic cell. Because this byproduct travels

from the postsynaptic cell to the presynaptic cell, which is the reverse direction of most synaptic communication, it has been labeled the *retrograde messenger*. There is now strong evidence indicating that the primary retrograde messenger involved in LTP is the gas nitric oxide (p.117).

Why is LTP believed to be the major cellular mechanism involved in learning and memory? The numerous reasons are too lengthy to discuss in detail. However, three fundamental properties of long-term potentiation have made it the primary candidate for a basic form of memory storage: *associativity*, *cooperativity*, and *input-specificity*. Associativity refers to the fact that both the presynaptic neuron and the postsynaptic neuron must fire in a carefully timed pairing if there is to be an enhancement of synaptic connectivity. For example, the role of the retrograde messenger is to alert the presynaptic cell that it should increase transmitter release. However, the messenger is only effective if the presynaptic neuron is already firing. In other words, for the connection between cell A and cell B to be strengthened, the two neurons must fire one right after the other. This associative property is fascinating because it is indicative of the long-held psychological understanding that two stimuli can become associated if they are paired together multiple times. The property of cooperativity refers to the cooperative involvement of multiple synaptic fibers, or afferents, involved in the firing of a neuron. This essentially means that a weak stimulation, one that fails to fire enough afferents, will not lead to the successful induction of LTP. Some researchers propose that this may be why insignificant stimuli are not remembered. Finally, the property of input-specificity suggests that in most cases only those neurons that fire during the learning experience, and not neighboring neurons, become connected.

After explaining how LTP works and why it is important, Squire and Kandel once again turn their focus to the psychological aspects of learning and memory. The final three chapters describe how priming, perceptual learning, and emotional learning occur; how skills and habits are developed; and finally, how biology can explain individuality. This final chapter is a strong reminder of the fact that humans are biological beings. Our perceptions, our actions, our philosophies, and everything about ourselves must therefore be traced to the activity of neurons in the brain. Though it is impossible to briefly review the tremendous amount of research presented by Squire and Kandel, these chapters are highly informative, presenting study after study in ways that are both interesting and enjoyable for readers with a limited scientific background.

In *Memory: From Mind to Molecules*, Eric Kandel and Larry Squire offer a unique view on the connection between learning and memory at the cellular level and higher cognitive processes such as our perception of the world. The authors' bold attempt to link molecular biology to a theory of human learning and memory storage makes *Memory* a valuable handbook for all social scientists, including applied linguists, interested in providing a neurobiological account of psychological phenomena.

PUBLICATIONS RECEIVED

Journals:

- Annual Review of Applied Linguistics*, 21, 2001
Applied Linguistics, 22, 2, June, 2001
ELT Journal, 55, 2, April, 2001
Journal of Multilingual and Multicultural Development, 21, 5, 2000
Journal of Multilingual and Multicultural Development, 21, 6, 2000
RELC Journal, 31, 2, December, 2000
TESOL Matters, 11, 2, March/April/May, 2001
TESOL Quarterly, 35, 1, Spring, 2001
Working Papers in Educational Linguistics, 16, 2, Fall, 2000

Books:

- Aitchison, J. (2001). *Language change: Progress or decay?* (3rd ed.). Cambridge: Cambridge University Press.
- Brumfit, C. (2001). *Individual freedom in language teaching*. Oxford: Oxford University Press.
- Canagarajah, A. S. (1999). *Resisting linguistic imperialism in English teaching*. Oxford: Oxford University Press.
- Chen, M.Y. (2000). *Tone sandhi: Patterns across Chinese dialects*. Cambridge: Cambridge University Press.
- Clark, E.V. (Ed.) (2000). *The proceedings of the thirtieth annual child language research forum*. Stanford: CSLI Publications.
- Coates, R. (1999). *Word structure*. London: Routledge.
- Dunayer, J. (2001). *Animal equality: language and liberation*. Derwood, MD: Ryce.
- Gardner, B. & Gardner, F. (2000). *Classroom English*. Oxford: Oxford University Press.
- Hadfield, J. & Hadfield, C. (2000). *Simple reading activities*. Oxford: Oxford University Press.
- Hadfield, J. & Hadfield, C. (2000). *Simple writing activities*. Oxford: Oxford University Press.
- Hollett, V. (2000). *Quick work: A short course in business English*. Oxford: Oxford University Press.
- Holm, J. (2000). *An introduction to pidgins and creoles*. Cambridge: Cambridge University Press.
- McNamara, T. (2000). *Language testing*. Oxford: Oxford University Press.
- Naunton, J. (2000). *Head for business*. Oxford: Oxford University Press.
- Palmer, F. R. (2001). *Mood and modality* (2nd ed.). Cambridge: Cambridge University Press.
- Peccei, J. S. (1999). *Child language* (2nd ed.). London: Routledge.
- Schendl, H. (2001). *Historical linguistics*. Oxford: Oxford University Press.
- Schulz, E., Krah, G. & Reuschel, W. (2000). *Standard Arabic: An elementary-intermediate course*. Cambridge: Cambridge University Press.
- Singh, R. (Ed.) (2000). *The yearbook of South Asian languages and linguistics*. New Delhi: Sage Publications.
- Stump, G. T. (2001). *Inflectional morphology: A theory of paradigm structure*. Cambridge: Cambridge University Press.

- Van Valin, Jr., R. (2001). *An introduction to syntax*. Cambridge: Cambridge University Press.
- Windeatt, S., Hardisty, D., & Eastment, D. (2000). *The internet*. Oxford: Oxford University Press.

Issues in Applied Linguistics SUBSCRIPTIONS

Volume 12, 2001☐ **Number 1** GENERAL ISSUE☐ **Number 2** GENERAL ISSUE**Volume 13, 2002**☐ **Number 1** GENERAL ISSUE: Topic to be announced☐ **Number 2** GENERAL ISSUE: Topic to be announced

Payment must be made by check or money order in U.S. funds. The following rates are per issue. Please double the amount if ordering a complete volume.

RATES: (per issue)☐ Student \$10.00☐ Individual \$18.00☐ Institution \$30.00

Shipping: Please include the appropriate fee and check one of the boxes below.

☐ Inside the US \$4.00 (per issue) ☐ Interational (surface) \$6.00 (per issue)☐ International Air Mail \$14.00 (per issue)

Amount Enclosed: \$ _____

(Make checks payable to **Issues in Applied Linguistics**. We regret that we cannot accommodate credit card orders at this time.)

Name _____

Address _____

Telephone/FAX _____

E-mail _____

Prepayment is required, so please mail your check and this form to:

Issues in Applied Linguistics

UCLA Applied Linguistics & TESL

3300 Rolfe Hall,

Los Angeles, CA 90095-1531

USA

Please direct subscription questions to <ial2@humnet.ucla.edu>

ial Back Issues

Did you miss an issue of *Issues in Applied Linguistics*? Update your collection by checking the boxes below and filling out the form on the back.

- ☐ **Volume 1, Number 1** Includes the following articles:
Towards a Critical Applied Linguistics for the 1990's by Alastair Pennycook
An Interview with Evelyn Hatch by Maria M. Egbert
- ☐ **Volume 1, Number 2** Includes general articles as well as:
SPECIAL FEATURE ROUNDTABLE:
Defining Our Field: Unity in Diversity
- ☐ **Volume 2, Number 1** Includes general articles as well as:
SPECIAL FEATURE ROUNDTABLE: *Language Education, Language Acquisition: Working Perspectives* by Four Applied Linguists
Leo van Lier, John Povey, Brian K. Lynch, and John Schuman
- ☐ **Volume 2, Number 2** SPECIAL ISSUE
Socialization Through Language and Interaction
Guest Editor, Elinor Ochs
- ☐ **Volume 3, Number 1** Includes general articles as well as:
SPECIAL FEATURE ROUNDTABLE:
Preparing Applied Linguists for the Future
- ☐ **Volume 3, Number 2** SPECIAL ISSUE
The Neurobiology of Language
Guest Editor, John Schumann
- ☐ **Volume 4, Number 1** GENERAL ISSUE
- ☐ **Volume 4, Number 2** SPECIAL ISSUE
Ethical Issues for Applying Linguistics
Guest Editors, Jeffrey Connor-Linton and Carolyn Temple Adger
- ☐ **Volume 5, Number 1** GENERAL ISSUE
- ☐ **Volume 5, Number 2** SPECIAL ISSUE
Applied Linguistics from an East Asian Perspective
- ☐ **Volume 6, Number 1** SPECIAL ISSUE
Sociolinguistics and Language Minorities

- ☐ **Volume 6, Number 2** SPECIAL ISSUE
Discourse Based Perspectives on Second Language Acquisition
- ☐ **Volume 7, Number 1** SPECIAL ISSUE
Proceedings from the First UCLA Conference on Language, Interaction, and Culture
- ☐ **Volume 7, Number 2** SPECIAL ISSUE
Applied Linguistics and Education
- ☐ **Volume 8, Number 1** SPECIAL ISSUE
Proceedings from the Second UCLA Conference on Language, Interaction, and Culture
- ☐ **Volume 8, Number 2** SPECIAL ISSUE
Proceedings from the Third UCLA Conference on Language, Interaction, and Culture
- ☐ **Volume 9, Number 1** GENERAL ISSUE
- ☐ **Volume 9, Number 2** SPECIAL ISSUE
The Organization of Participation
- ☐ **Volume 10, Number 1** GENERAL ISSUE
- ☐ **Volume 10, Number 2** GENERAL ISSUE
- ☐ **Volume 11, Number 1** SPECIAL ISSUE
Nonnative Discourse
- ☐ **Volume 11, Number 2** GENERAL ISSUE

back issues order form

Please mark on the front of this form which issues you would like to purchase.
The following rates are per issue.

RATES: (per issue)

☐ Student \$10.00 ☐ Individual \$18.00 ☐ Institution \$30.00

Shipping: Please include the appropriate fee and check one of the boxes below.

☐ Inside the US \$4.00 (per issue) ☐ Interational (surface) \$6.00 (per issue)
☐ International Air Mail \$14.00 (per issue)

Amount Enclosed: \$ _____

(Make checks payable to **Issues in Applied Linguistics**. We regret that we cannot accommodate credit card orders at this time.)

Name _____

Address _____

Telephone/FAX _____

E-mail _____

Prepayment is required, so please mail your check and this form to:

Issues in Applied Linguistics
UCLA Applied Linguistics & TESL
3300 Rolfe Hall,
Los Angeles, CA 90095-1531
USA

Please direct subscription questions to <ial2@humnet.ucla.edu>

\$20 service charge for all returned checks.

Information for Contributors

ial is a refereed scholarly journal published by the graduate students of the UCLA Department of Applied Linguistics & TESL. Our aim is to publish outstanding research from students, faculty, and independent researchers in the broad areas of discourse analysis, sociolinguistics, language acquisition, language analysis, language assessment, language education, language use, and research methodology. We are particularly interested in publishing new departures and cross-disciplinary applied linguistic research.

Contributions, including book reviews, interviews, and articles, should be submitted in three copies and sent to Editor, *Issues in Applied Linguistics*, Applied Linguistics & TESL, UCLA, 3300 Rolfe Hall, P. O. Box 951531, Los Angeles, CA 90095-1531.

Manuscripts must be printed double spaced (including the references, examples, etc.) on one side only of white 8^{1/2} x 11 or A4 paper and must use black printing ink. Please attach a cover sheet with the following information: a) the full title of your paper, b) your name, c) name and address of your institution (or other address where you can best be reached), d) your phone number (work and home), FAX number and E-mail address (if any), and e) short title of your paper (for running head). The first page must include an abstract of the article which is less than 150 words. Authors should also supply a biographical note of approximately 50 words.

Essential notes should be indicated by superscript numbers and printed at the end of the article (but before the references).

References and citations should be made using APA Format: for example, (Cazden, 1988, p. 53), or (Sinclair & Coulthard, 1975). A series of references should be listed in alphabetical order: for example, (Cazden, 1988; Goodlad, 1984, Lemke, 1991), or (Long, 1996; Long & Porter, 1985). A series of references for the same author or authors should be ordered in chronological order as follows: (Hall, 1993, 1995, 1997). Use letters a, b, c to distinguish citations of different works by the same author in the same year. All references cited in the text must be listed in the references list alphabetically. For more information, please see the *Publication Manual of the American Psychological Association, Fourth Edition*.

Submission of an article or book review is taken to imply that the manuscript has not been previously published or submitted simultaneously to another publication. If a related article has been published elsewhere, this should be noted. Acceptance of articles will be contingent on the assignment of copyrights to the UC Regents, the publisher of *ial*.

CALL FOR PAPERS

ial is a refereed journal published by the graduate students of UCLA's Department of Applied Linguistics & TESL. We are particularly interested in publishing new departures and cross-disciplinary applied linguistic research. In the past we have published quality work on all of the following topics:

Discourse Analysis
Functional Grammar
Language Socialization
Bilingual Education
Sociolinguistics
Professional Ethics

Language and the Brain
Second Language Acquisition
Language Education
Research Methodology
Conversation Analysis
Language Minorities

ial has also published book reviews and interviews with notable applied linguistics scholars. To submit your manuscript, please send three copies to the following address:

Editor

Issues in Applied Linguistics

UCLA Applied Linguistics & TESL

Box 951531, 3300 Rolfe Hall

Los Angeles, CA 90095-1531

E-mail: ial@humnet.ucla.edu

For more information please visit our website at:

<http://www.humnet.ucla.edu/humnet/teslal/ial>

ANNOUNCEMENT: The Evelyn Hatch Award Fund

Moneys from the Evelyn Hatch Award Fund at UCLA are now used to recognize and reward the efforts of the editor(s) of *Issues in Applied Linguistics* with a modest annual honorarium.

If you would like to support these efforts, please make a tax-deductible contribution to the Hatch Award Fund. Send your check or money order made out to "UC Regents" and ear-marked for the Evelyn Hatch Award to:

Ms. Mila August, MSO

UCLA Applied Linguistics & TESL

Los Angeles, CA 90095-1531

If you have any questions about the Award, you can contact Professor Marianne Celce-Murcia at the above address or on e-mail at:

celce-m@humnet.ucla.edu.

BOOK REVIEWS

Language Testing

Tim McNamara

Reviewed by Kirby J. Cook

How Children Learn the Meanings of Words

Paul Bloom

Reviewed by Masahiko Minami

Memory: From Mind to Molecules

Larry Squire and Eric Kandel

Reviewed by Sheila Crowell